NEW PERSPECTIVES ON THE HISTORY OF COGNITIVE SCIENCE

Neurocognitive Development and Impairments

Series Editor
Csaba Pléh

NEW PERSPECTIVES ON THE HISTORY OF COGNITIVE SCIENCE

edited by
Lilia Gurova, László Ropolyi and Csaba Pléh

AKADÉMIAI KIADÓ, BUDAPEST

ISBN 978 963 05 8797 6
HU ISSN 1786-2620

Published by Akadémiai Kiadó
Member of Wolters Kluwer Group
P.O. Box 245, H-1519 Budapest, Hungary
www.akademiaikiado.hu

Printed in Hungary

TABLE OF CONTENTS

CONTRIBUTORS

Andrew Brook	Carleton University, Ottawa, ON, Canada andrew_brook@carleton.ca
Bálint Forgács	Budapest University of Technology and Economics (BME), Budapest, Hungary forgacsb@cogsci.bme.hu
Lilia Gurova	New Bulgarian University, Sofia, Bulgaria lilia.gurova@gmail.com
Lena Kästner	Ruhr University Bochum, Bochum, Germany mail@lenakaestner.de
Olga Markič	University of Ljubljana, Ljubljana, Slovenia olga.markic@guest.arnes.si
Leone Montagnini	University "Federico II", Naples, Italy leone.montagnini@unina.it
Gabriele M. Mras	University of Vienna, Vienna, Austria gabriele.mras@univie.ac.at gabriele.mras@wu-wien.ac.at
Csaba Pléh	Eszterházy College, Eger, Hungary pleh.csaba@ektf.hu
Andreas Reichelt	University of Vienna, Vienna, Austria andreas.franz.reichelt@univie.ac.at
Nicole Rossmanith	University of Vienna, Vienna, Austria EMAIL
Jean-Michel Roy	University of Lyon, Lyon, France Jean-Michel.Roy@ens-lyon.fr
Hayo Siemsen	INK, FH Emden/Leer, Emden, Germany hayo.siemsen@gmail.com
Filomena de Sousa	Technical University of Lisbon Lisbon, Portugal, mfisousa@fc.ul.pt
Sven Walter	University of Osnabrück, Osnabrück, Germany s.walter@philosophy-online.de
Maximilian Wontorra	University of Leipzig, Leipzig, Germany wontorra@rz.uni-leipzig.de

FOREWORD

This book is the several times revised and edited version of a small section held on the history of cognitive science at the XXIII International Congress of History of Science and Technology held in Budapest, between July 27 and August 2, 2009. Philosophers, historians, and active cognitive scientists took part in the symposium.

It is our hope that the small volume is able to reconstruct the openness and lively interchange of ideas that characterized the meeting. The editors believe that the emerging new discipline of cognitive science is a very apt domain to test the openness and maturity of the historians of science.

We would like to thank the publisher and Lőrinc Vajda for the welcome of our project, and the careful editorial work.

The Editors

EXISTING AND WOULD-BE ACCOUNTS OF THE HISTORY OF COGNITIVE SCIENCE: AN INTRODUCTION[1]

Csaba Pléh and Lilia Gurova

> *Science [...] is the source of a deep tension in the modern world. On the one hand, we have been formed by the traditions of our culture. [...] On the other hand, modern empirical sciences have reshaped our world and the way this whole world is interpreted. (Hans-Georg Gadamer 1983: 209)*

In this introductory chapter, the authors have three aims:

1) To summarize the accepted vision of the history of cognitive science, and somehow characterize the different weights put on different historical and substantial parts of the story, according to the sometimes diverging constituting disciplines.
2) To show that in (especially continental) European philosophy and psychology, there is an alternative prehistory of cognitive science, along the lines introduced by Brook (this volume).
3) Finally to indicate how some of the recent developments can be seen as replies to the traditional challenges regarding the supposedly reductionist theories of the mind.

Different approaches to the history of cognitive science

Modern, late 20th-century cognitive science is usually seen from one of two perspectives: either from the point of view of psychology's internal development, or from the point of view of the interaction between the nascent computing science and the different human and social sciences dealing with man as the knower.

The essence of the traditional image is that Cognitive Science (from now on CogSci) took shape as an integration of the special sciences dealing with the different aspects of mainly human cognition, and the upcoming computing technology and computer sciences. This approach treats cognition in a neutral way that is referred to as *the computational theory of the mind,* and assumes that the essence of knowledge is undivided: it is basically propositional, and its underlying architecture is based on symbol manipulation according to rule-based steps.

We show the formation of what has become called classical CogSci from three main perspectives: that of psychology, computer science, and philosophy.

[1] The authors benefited from several discussions regarding this outline. Csaba Pléh was encouraged by discussions with his Middle European Cognitive Science master students coming from Vienna, Silvia Maier, Elisabeth English, and Fabian Simank, and his PhD students, Zsombor Várnagy, and especially Bálint Forgács. In preparing the final version of this chapter and the entire book Csaba Pléh was enjoying a scholarship at the Collegium de Lyon, ENS Lyon and a support within the framework of the TÁMOP-4.2.2.C-11/1/KONV-2012-0008 (Social Renewal Operative Program) project titled *The application of ICT in learning and knowledge acquisition: Research and Training Program Development in Human Performance Technology.* Said project was implemented by the support of the European Union and the co-financing of the European Social Fund.

The psychologist's perspective

As described in several autobiographic accounts (Bruner 1983, 1997; Miller 2003), and in overviews such as Gardner's (1985) by now classic personalized treatment, for a generation of psychologists maturing in the 1950s and 1960s, a challenging and non-trivial shift appeared in the form of *the cognitive revolution.* Through a variety of different influences, they came to realize two important aspects: the human mind has a real existence, and there are scientific ways to study the mind, without loosing the objectivity professed by behaviorism and the governing positivist philosophy. Though for most of the participants the cognitive revolution seemed to be a radical breakaway from behaviorism, reflection a generation (or even two) later, as Roy and others show in the present volume, suggests that several similarities between behaviorism and CogSci are realized today such as the hope for a reductionist account of the mind, and a neglect of the use of phenomenological introspection.

From the perspective of modern (American) psychology, the emergence of cognitive science came in two waves: first cognitive psychology was born, and then with a further step towards abstraction on the one hand, and towards natural science on the other hand, cognitive science was born too, as a result of a second move or revolution, if you like.

These two stages are illustrated in Figure 1.

In the first stage, a vision was gradually created according to which *man can be seen as a being actively modeling the environment,* and human behavior can only be understood with reference to these models. That was the typical approach taken in the early 1960s by the Harvard Center for Cognitive Studies (see Bruner, Oliver, and Greenfield 1966; and the collection of papers of Bruner 1973), but it appeared in the popular book of George Miller (1962) too, which was among the first to reclaim the study of 'mental life' for psychology, a turn that can be interpreted historically as a *return* to the American mentalism *à la* William James (see Brook, this volume).

Table 1 indicates some of the converging influences that were encouraging as well as motivating the first generation of cognitivists within psychology to break away from the behaviorist credo.

This classical cognitive psychology was mainly using reaction time data and confusion errors to represent the modeling aspect of the human mind with its different models of information processing and mental representation. A problem inherent to this approach was the representativeness fallacy. Psychologists in the 1960s had assumed that letter recognition, for

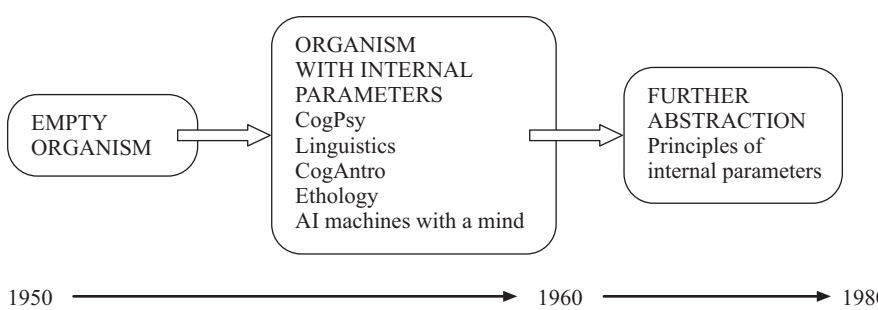


Figure 1. The two stages of the birth of modern Cognitive Science as seen by psychologists

example, and other complex tasks are good models for the mechanisms of the entire human mind. By studying memory for artificial letter sequences you may learn how memory works in general. The approach has now a half a century history, and in this time it has become the victorious one within modern psychology, what is not based only on self-praise among the cognitivists, but on a few Nobel prizes like H. Simon (1978) and D. Kahnemann (2002), both in economy, and R. Sperry (1981) in medicine and physiology.

Table 1. Important factors in forming the cognitive approach within psychology as a modeling approach (after Pléh 1998)

Field	Impact, incentive
Behaviorism	Objectivity, input–output analysis
Information theory	Quantitative description, communication model
Cybernetics	Regulation, feedback, learning machines
Mathematics	Algebraic structures, axiomatization
Computer science	Machine analogy, flow charts
Linguistics	Mentalism, structures, and rules
Ethology	Innate species-specific modeling, releasers
Radiotechnology	Information processing, coding

The importance of the cognitive movement to the entire field of psychology is also shown by the general attention of the psychology profession towards its cognitive trends. According to the multi-method study of Haggbloom et al. (2002) that looked for the most influential psychologists of the 20th century using journal citations, textbook references, and expert judgments, out of the 100 hallmark psychologists, 21 selected were classified by us as cognitive psychologists from the contemporary scene. This is shown in Table 2, even excluding further

Table 2. Cognitive psychologists from the list of the 100 most influential psychologists of the 20th century, in descending, and left to right order (from the data of Haggbloom et al. 2002)

1. Piaget, J.	8. Simon, H.	15. Posner, A.
2. Hebb, D.O.	9. Chomsky, N.	16. Loftus, E
3. Miller G.	10. Osgood, C.	17. Lurija, A.
4. Bruner , J.	11. Bower, G.	18. Gibson, J.
5. Neisser, U.	12. Sperry, R.	19. Gibson, E.
6. Brown, R.	13. Broadbent, D.	20. Rumelhart, D.
7. Tulving, E.	14. Shepard,R.	21. Tversky, A.

cognitively relevant authors like Köhler, who were not very active in the 1960s, which was taken as cutpoint for 'contemporariness'.

The *cognitive trend* showed up in other sciences as well. Rather directly in anthropology and sociology, even in the form of self labeling: many authors started to view society and the differences between societies as a modeling problem (see Cicourel [1974] for cognitive sociology, and the summary of D'Andrade [1995] that shows the two decades of his efforts towards a cognitive anthropology, as well as Colby, Fernandez, and Kronenfeld 1981; and Dougherty 1985 for anthologies of cognitive anthropology.) Nevertheless, the inner shift was true for most of the general theories of ethology as well (Csányi 1988).

Table 3 shows an overview of the cognitive approaches in the different special social sciences. This list is by far not exhaustive – just think of cognitive neuroscience that has taken shape in the 1990s, and of course the different branches of computer science.

Table 3. The cognitive branches of some specific social sciences (after Pléh 1998)

Domain	Leading topic	Key concepts
Cognitive psychology	Information processing	Representation, coding, stores
Cognitive anthropology	Cultures as models	Classification, relativism
Cognitive sociology	Social representation	Rules, classification
Cognitive linguistics	Language and thought	Metaphors, coherence
Cognitive ethology	Species-dependent modeling	Releasers, behavioral model

Within all these cognitive branches, including cognitive psychology, there are two basic underlying topics:

1) *Modeling the world* – in the form of representations in humans. It is crucial to the essence of higher forms of life.
2) *External stimuli can determine behavior only through interpretation by the organism,* or by the mental representation of culture.

Central representation becomes more and more sophisticated in the new cognitive theories, as illustrated in Figure 2.

This new cognitive psychology has gradually become connected to other more formal disciplines dealing with knowledge, such as epistemology, logic, and most of all the new inspiration and excitement of the 1960s which *came from computer science and machines.* In a way, as we shall discuss it later on, cognitive psychology and later CogSci could even be seen as a naturalistic interpretation of epistemology. The relations and interactions between psychology and CogSci, especially the discussion between psychology, philosophy, and the machine expertise – computer science, AI, machine vision, language parsing – has culminated in the 1970s in the *'second cognitive revolution',* which led to a cultivation of questions on a one-step-higher level of abstraction.

It is no exaggeration to talk here of a second revolution. As the book edited by Johnson and Erneling (1997) shows sometimes mockingly, in a way cognitivists are always in a need of

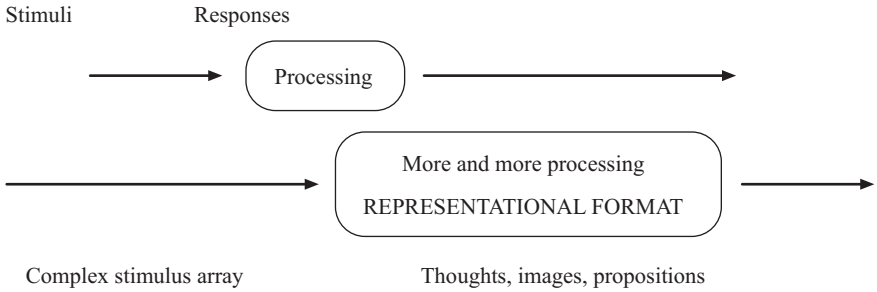


Figure 2. The change from a behaviorist determination towards cognitive determination with more internal processing

newer and newer revolutions. The mainstream cognitive psychologists already made a courageous move towards a neutral characterization of cognition, abstracting from the biological, neural mechanisms responsible for it (e.g., Broadbent 1958). Cognitive science in its first stages, even within cognitive psychology, deliberately neglected the brain-based interpretation of its models. In this regard again, it was unlike the radical descriptive behaviorism of Skinner (1953). Early cognitive psychologists even deliberately moved away from the rather abstract physiological interpretation of early initiators like Donald Hebb (1948), who coined the term 'Conceptual Nervous System' for the non-specified use of CNS thinking he patronized. The cognitivists also consciously put on the side the evolutionary issue of what types of beings are responsible for cognition. The ideal level of abstraction for classical CogSci intended to be a common denominator between machines, humans, animals, and whatsoever that can be thought of entertaining ideas. Cognitive scientists are talking metaphorically about Coke bottles, or the inhabitants of China as representing cognitive agents, and one of the proponents of the neutral functionalism of early CogSci talks even of angels as possible cognitive agents. Jerry Fodor (1975) in his Language of thought openly entertains the idea, even if only on the level of thought experiments, that angels might have some idealized cognitive processes.

This kind of physical neutralization was important because cognition for this school was – in contrast to present-day embodied slogans – deliberately disembodied. This attitude enacted a functionalism that – while it was rejuvenating the Aristotelian approach – basically claimed for a substance neutralism. The very same way as we can characterize the work of early computing machines in the form of flowcharts for example, without being interested in issues of electric circuitry (i.e., the hardware), we can do the same thing with human thought (Putnam 1960). The first of the two-volume excellent reader edited by Ned Block (1980) on the philosophy of psychology – very tellingly not yet on CogSci![2] – deals with the issues of functionalism in detail. Block characterized the two poles as either being too liberal, or too psycho-chauvinist. The solution offered by Block was to claim that there are certain types of mental computations that could be formed by many other entities beside human brains. This attitude still functions as a point of reference for a rationalistic–mentalistic–functionalist phi-

[2] It is very telling that about a decade later readers of similar scope, such as the one edited by Goldman (1993), call themselves cognitive science and philosophy readers. The metonymy has taken place: psychology is replaced by CogSci.

losophy of mind. According to this approach, it does not matter what type of being is doing the computations responsible for cognition.

From cognitive psychology to cognitive science

One technical aspect in the late 1970s, when the special cognitive sciences turned into cognitive science at large, was the promise to understand computations by more explicit models brought in from sophisticated AI, such as the approach of David Marr (1982).

As for the *dating of these moves,* i.e., dating the 'second cognitive revolution' there are several accounts. One, entertained sometimes by George Miller (2003) as well, classifies the immediate prehistory of cognitive science as being a part of cognitive science itself. Miller lists many AI conferences, the important book by Bruner, Goodnow, and Austin (1956), but also gives a special moment of inception, as well as the birth date, of CogSci. According to this account, the moment would date back to a conference in the fall of 1956, on September 11, the second day at a "symposium organized by the 'Special Interest Group in Information Theory' at the Massachusetts Institute of Technology" (Miller 2003: 142). Here, the new cognitive psychologists, such as Miller himself, the linguists, represented by Chomsky, and the researchers of the new computer approaches to thought, represented by Newell and Simon's (1972) problem-solving programs, have met. This account is shared by Gardner (1985), and Smith (1990), based on interviews with the participants.

Cybernetics and CogSci: An excursion

There is an interesting recent controversy regarding *the importance and relevance of early cybernetics to the birth of CogSci.* The French soft-Heideggerian philosopher Jean-Pierre Dupuy in a book (which was published in French in 1994, and in English in 2000, and its rewritten version was published in 2009) has a twofold perspective. His first point is that traditional treaties of the history of CogSci, as well as the everyday practice of CogSci disregard the cybernetic past. This is, incidentally, philologically incorrect. The majestic two-volume book of Margaret Boden (2006) concentrates on the cybernetic prehistory, and even in accounts of advocates of a 'paradigm shift' in the sense of Kuhn (1970) during the 1960s towards cognition, as of G. Miller (2003), the importance of the information theoretical and cybernetic past is clearly seen. This was even practiced in his own work by George Miller, not only in his classic papers about the limitations of information uptake, but also in a general theory about feedback mechanism between internal plans and their execution (Miller, Galanter, and Pribram 1960).

According to Dupuy, the cybernetic past represented by people like Norbert Wiener (1948, 1950), and William Ross Ashby (1956) is ignored by mainstream CogSci. His other message is that if the cybernetic past is entertained, it is misunderstood in a techno jargon. The real message of cybernetics has to do with the *intentionality issue.* There is a human message of cybernetics that cannot be reduced to information processing machines.

As for his first claim, Dupuy argues that 2-3 decades before the advent of the cognitive movement in the late 1970s the idea of a computational theory of mind was formulated by

the cyberneticians. With the idea of regulation and with the theoretical promise of a thinking machine, Wiener and his followers were already proposing that thinking was nothing but computation. That image has been proposed by Wiener (1950) as a social utopia as well.

The criticism of Dupuy is sometimes misguided. He claims for example that the early McCulloch-Pitts (1943) model of neuronal nets realizing a logical calculus was ignored by the cognitive people. In reality, in the mid-20th century the neuronal net idea was taken very seriously. Marvin Minsky and Papert (1969) in their *Perceptron* model criticized all single-layered neuronal nets as having insufficient computational power. The later non-representationist approach proposed by the connectionist modeling CogSci people took the neuronal net idea very seriously, but again proposed multi-layered nets, and massive parallel processing in neuronal nets (Rumelhart, and McClelland 1986; McClelland, and Rumelhart 1986). Neural nets are taken seriously by present-day neural modeling, too, but with the emphasis on hidden layers between the observable behavior and the stimulus array. Later synthesizing works by Arbib, Érdi, and Szentágothai (1997) also started from the ideas represented by the McCulloch-Pitts model when trying to relate neural circuitry and cognitive modeling.

There is an older historical aspect as well. Piaget (1963) in his attempt to find a neural basis for his logical calculus ideas saw from very early on the revolutionary importance of the McCulloch-Pitts model as a way to embody logics in a neural net. Of course, we have become smarter in half a century. If a binary logical calculus with and/or gates can be implemented in the neural nets of snails, then what is the advantage of the enormous brain of humans? Thus, it is a separate issue to claim that we have a material system that *could implement* a logical calculus, and another separate issue whether this system can also realize the calculus on the level of open behavior and metacognition. That is certainly not true of the snail neuronal nets.

Dupuy in his Heideggerian fever at the same time treats the cybernetic movement as well as present-day CogSci as being too materialistic. For Dupuy, the basic issue with computational models of thought processes is the embodiment in physical hardware. However, for many cognitive scientists, the essential moment of the computational approach is not the physical embodiment but the central role of form in processing. *The form of thought* is the central aspect of an ontologically neutral cognitivism. That approach – in the articulation of a syntactic theory of mind, or in other representational formats like the image-based models of cognition, has centered on the notions of information and representation.

Dupuy is certainly right in showing that issues of function and teleology cannot be excluded from machine-based metaphors of CogSci. But this is the approach taken by people who highlight that even representations in the cognitive sense have to be interpreted in a semantic way. The mind is not only a syntactic machine in the abstract sense but an intentional semantic machine as well (Dennett 1987).

Institutional aspects of establishing CogSci

A more conservative approach that is more in line with the idea of two stages in the birth of CogSci dates the (American) birth of CogSci proper to a Sloan Foundation initiative that was supporting the articulation of the new science from the mid-1970s on. We tend to sympathize with this latter attitude. Though all events in 1956 were important because of the continuity of personal life events (i.e., the protagonists including Chomsky and Miller, were

very young then, 28 and 36, respectively), the formation of a new discipline usually entails some institutional aspects as well. Psychologists are familiar from the history of their own discipline that shaping a new discipline entails several institutional aspects (Danziger 1990; Kusch 1995; Pléh 2008). The institutional changes involve the following (extended with an indication when this happened for CogSci):

1) Creation of Associations in 1979.
2) Establishing of journals between 1972 and 1976.
3) Creation of textbook-like reference works between 1981 and 1990.
4) Creation of symbolic places, centers, departments in 1986.

Cognitive science as a new disciplinary movement also meant conferences. It also meant the founding of a new society, *The Cognitive Science Society,* that was created during the first CogSci conference in La Jolla (CA), in 1979. The journal *Cognitive Science* started in 1977.

Cognition, which was started in 1972was the first interdisciplinary journal of the field initiated by two MIT–Harvard graduates, Tom Bever and Jacques Mehler. Until the early 2000, Mehler was the single editor of this extremely influential journal, which has by now 12 issues annually. Mehler and Franck (1995) give a detailed account of the intellectual atmosphere leading to the establishment of the journal, namely the need to have a journal intellectually open to many theories, and to many disciplines. In reality, however, *Cognition* as well as other journals of cognition at large became dominated by psychologists, and by topics of individual laboratory-based cognitive research. During this period within psychology, other journals also turned into cognitive ones, like the *Journal of Verbal Learning and Verbal Behavior* which became *Memory and Language.*

From the 1980s on, several textbook- or handbook-like publications appeared. Norman (1981), Posner (1989), and Osherson (1990) more or less served the important conceptual and substantial integration that is needed for the socialization of the would-be followers of a serious new discipline.

The establishment of cognitive science also entailed *the creation of university departments or centers.* Interestingly enough, but not surprisingly, knowing the cautious moves of university administrators, this step took the longest. At UC San Diego, the administrative unit was created in 1986, as a secession from Psychology, after a decade long cognitive involvement of people there like D. Norman, D. Rumelhart, G. Mandler, the late Elisabeth Bates, and the president of the university, Richard Atkinson, both a re-known memory model researcher and a successful university administrator. At MIT, the *Department of Psychology,* established by an excellent neuroscientist H. L. Teuber, was turned into the *Department of Brain and Cognitive Science* in 1986 (http://bcs.mit.edu/aboutbcs/history.html). As an expression of the importance of *science* in the name of the field and the department, it moved to the Science faculty at MIT. In several places, the cognitive institutional units form more of an umbrella organization. This is happening at places such as the University of Colorado, Boulder with *The Institute for Cognitive Science* (http://ics.colorado.edu/people/associates.html), or in Rutgers University (http://ruccs.rutgers.edu/ruccs/index.php), where a center with a small direct faculty is coordinating the cognitive aspects of the work of people in different departments. This is the model of Ohio State University, and Indiana University, Bloomington as well.

Today, there are a few dozen departments and a dozen centers or institutes of cognitive science, sometimes in combinations expressed in their names as well, like Cognitive Science and Linguistics, Brain and Cognitive Science, Cognitive Science and Psychology, and the like. It is not our aim to survey the entire field in a historical introductory paper. The interesting point is to show that the different arrangements imply different attitudes. Again, looking back at the history of psychology, it is rather telling that in the 1980s and 1990s several psychologists moved to CogSci, emphasizing the science aspect of their own discipline, moving their institutions from social science to the science faculty. The division and hesitation of psychology between philosophy, social science and the humanities branches is with us since the 1870s. What is very telling about the general natural science move of psychology in the late 20th century is the fact that in the 1950s and early 1960s the division was institutionally surfacing at several American East coast schools by establishing departments of Social Relations like those at Harvard and Columbia, and leaving psychology for the hard-headed behaviorists. That was the time when part of psychology was flirting to associate itself with sociology and anthropology. The CogSci trend shows the opposite move: the other, naturalistic part of psychology wants to associate with math, computation, and biology.

Interestingly enough, from 2009 on, the Cognitive Science society also publishes a new digest journal that carries reviews under the name of *Topics in Cognitive Science.* (It might be related to the great success of an independent journal of the *Trends* group, that of *TICS,* i.e., *Trends in Cognitive Sciences*). The *Topics* journal in 2010 presented a survey of the 30 years of CogSci (Barsalou 2010). Barsalou gives us a comprehensive picture, and all the authors of the volume try to give a specialized look back at the scene from the angles of philosophy, linguistics, anthropology, AI, and education.

CogSci journals become dominated by psychologists

Out of the analyses given in the new journal, the most telling for our perspective taken here is the paper by Gentner (2010). She shows the same bias we are indicating here: while 30 years ago CogSci started with the promise of becoming a comprehensive integrative field, it has become more and more a field invaded by laboratory psychologists. Figure 3 shows Gentner's data in a simplified form.

Incidentally, it is worthy to indicate towards later developments that this was by far not an unavoidable fate neither in journal policy, nor in intellectual outlook. Another cognitively

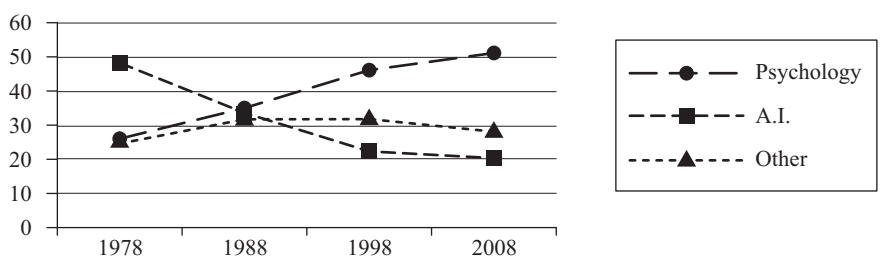


Figure 3. The increasing dominance of psychology within cognitive science papers in the journal *Cognitive Science* (data in percentages from Gentner 2010)

oriented multidisciplinary journal, *Behavioral and Brain Sciences* (notice the missing word cognitive here), which was launched around the same time in 1978 by Stevan Harnad, had the aim to be interdisciplinary with a more biological flavor. (Not surprisingly, since Harnad was the last PhD student of Donald Hebb.)

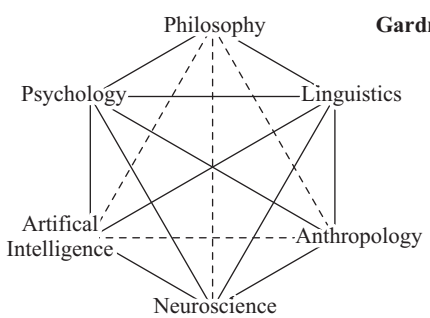
The original scope of the BBS journal was intended in the following way by Harnad (1978).

1) *Behavioral biology* (including behavior genetics, animal communication and intelligence, human ethology, invertebrate, lower vertebrate and mammalian behavior, primatology, sociobiology, etc.).
2) *Cognitive science* (including artificial intelligence, human information processing, linguistics, mathematical models, philosophy and philosophy of science, psycholinguistics, psychophysics, etc.).
3) *Neuroscience* (including higher CNS function, invertebrate neurobiology, human neuropsychology, motor systems, neuroanatomy, neuroethology, neurochemistry and neuropharmacology, sensory systems, etc.).
4) *Psychology* (including clinical, cognitive, comparative. developmental, personality, social and physiological psychology, experimental analysis of behavior, etc.).

More importantly, in BBS this variety of outlook remained there for over 30 years; it is there even today. The four target papers in 2010 are on fertility, demography, philosophy, and psychology. The balance is clear in the selection of commentaries as well. There are 33 commentators coming from psychology departments, 12 coming from cognitive science and related interdisciplinary units (centers, departments and the like), 10 from departments of philosophy, and 5 from neuroscience- and biology-related units.

The intellectual status of CogSci

As also presented in our volume in papers by Reichelt and Rossmanith, Roy, and others, CogSci has different proposals regarding its intellectual structure. Figures 4 to 6 show graphically the different existing interpretations of how cognitive science is to be represented in the system of sciences, and what is its interdisciplinary role. Figure 4 is the Sloan Foundation



Gardner's Hexagon

Connections among the Cognitive Sciences
Key: Unbroken lines – strong interdisciplinary ties
Broken lines – weak interdisciplinary ties

Figure 4. Cognitive science as the summary of cognitive visions in the different disciplines (after Gardner 1985; taken up by Miller 2003)

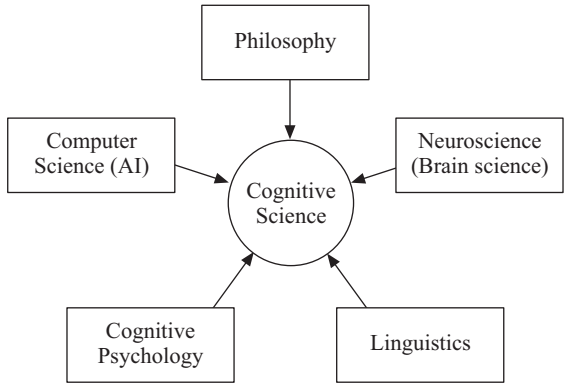


Figure 5. Cognitive science as a crosstalk of disciplines
(national Chinese University in Taiwan http://www.ncku.edu.tw/~iocs/en_US/faculties/people.php)

report approach made famous by Gardner (1985), which corresponds to a recognition of the internal aspects of each participating discipline, and tries to relate them with closer (solid lines) or farther (broken lines) interpretive or reductionist connections. Each discipline keeps its autonomy – CogSci is 'merely' expressing the joint, or common topics. This notion corresponds to the switchboard-type institutes or centers of cognitive science.

Figure 5 shows CogSci as an integrative discipline that takes data and cognitive approaches from all the other fields, but constitutes itself as an independent field. CogSci corresponds here to the idea of a separate department with many professional competences but with its own agenda: CogSci is the merger domain of all the relevant fields. Incidentally, that was the image Piaget (1970) has given to the system of sciences, but for him in the center field place of CogSci one finds, well, psychology.

Finally, Figure 6 shows an image proposed by Pléh (1998). In this representation CogSci is the substantial and attitude-like intersection of the different composing fields. It is a collec-

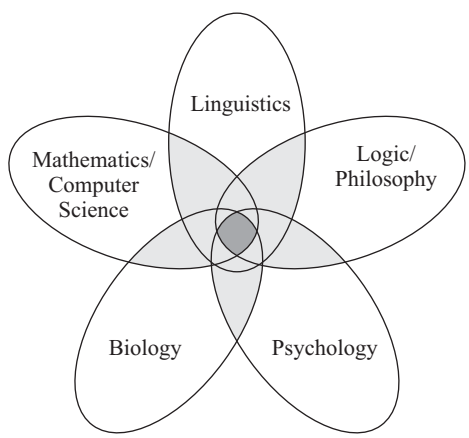


Figure 6. Cognitive science as a crossover of the disciplines (Pléh 1998)

tion of issues – such as brain and mind relation, empirism and innatism in development, the form of thought, etc. – which are important to all the neighboring fields. (Incidentally, this is the image most preferred by students of Csaba Pléh.)

The intersection, the center field ball, can change size with the development of the field, and it can also move towards one of the corners by advocates of CogSci coming from different original disciplines. This is very well seen in the textbooks or handbooks. The Oxford analytic philosopher of science, Rom Harré (2002), who is now at the London school of Economics naturally moves the ball towards categorical issues of philosophy. Early enthusiastic followers of the marriage between strong AI and CogSci moved it towards computer modeling (Norman 1981). Posner (1989), one of the founders of modern experimental cognitive psychology, naturally has a psychological bias. The linguist Jackendoff (1992, 2009) has a natural tendency to concentrate on issues of language as the crucial themes in understanding the workings of the mind. This is not a necessity, however. The linguistic philosopher Harnish (2002) has a tendency towards interpreting general computational models. By far the most balanced treatment is the four-volume handbook edited by Osherson et al. (1990). However, it is the hardest to read because lacking a bias also means lacking a perspective.

Classical cognitive science as a syntactic theory of mind

The very idea to make the processing of form (or "shape", as Fodor [1985] prefers to call it) central in information processing existed well before the advent of the machine metaphor, and the impact of information theory (Broadbent 1958). It was present in modern structural linguistics, and in the philosophy of language as well. The "discovery" was not the great leap provided by machines processing information, and the application of the ideas to humans, but the very realization itself: the existential proof that man can be interpreted in information processing terms, the very fact that devices that follow these principles can do their work (beside Broadbent, see a later synthetic work within psychology: Lachman, Lachman, and Butterfield 1979).

This attitude is usually referred to as the *symbol processing metaphor of human cognition* that was the general unifying idea of CogSci at its inception in the late 1970s. This was the metatheory that we are likely to call today Classical CogSci. Walter (this volume) even goes as far as to characterize this as Classicism. It is easy to summarize its basic tenets (Newell 1980, 1989; Newell et al. 1989).

1) Human cognition can be characterized as a recoding process of several steps working over symbols. It turns representations into different formats, the final format being a propositional calculus.
2) Human information processing defines a machine that works in a sequential, linear manner.
3) Our processing limitations are of a single common kind because all cognitive requiring processing resources are translated into the language of a joint resource.
4) Processing requires the cooperation of relatively small capacity operative storage systems, and large capacity background stores. Our knowledge is stored in the background memories, while operative memory represents the activated knowledge, and incoming input.

5) Our cognition has but a single active processing unit that would correspond to the Cartesian unity of thought, and to the CPU of a classical computer with a Neumann-architecture.

It is a challenging aspect of this approach to raise the question: *In what sense does thought itself have a unified structure?* The classical CogSci was mainly dealing with relatively slow processes, and it was not too much interested in the peculiarities of perception, and the social embedding of thought. How far is propositional organization a feature of the theory of mind itself? Or is it only a descriptive convention? The language of thought, according to the LOT hypothesis promoted by Fodor (1975), rephrased the Leibnizian topic in the light of the machine age. Programming languages translate instructions of a higher order into instructions of a lower order, to arrive at the end at a machine code. Similarly in humans: there is a final language, *a mentalese.* Human thought can be interpreted as an organization, where some final instances (the very propositional organization) correspond to a pre-wired language.

The proposal that the language of thought is similar to a machine code, is intended to avoid the infinite regress. While all thought is symbol manipulation, i.e., translation, finally there is a language, a form of thought provided by nature herself. The LOT is a linguistic a priori system.

Altogether, in classical CogSci we arrive at a rather peculiar view of human life. We have a thoughtful contemplative man, who concentrates on the form of representations (it is even questionable whether the LOT has an internal semantics, or it is a purely syntactic engine). The ideal is pure cognition, undisturbed by episodic representations, and it is achieved by using unified principles. This is a rather dry and language-centered vision of the human mind. We have to remember, however, that this style of dryness is familiar from old times. It is not due to machines but to the cognitive bias of the rationalists. If all of this seems to be empty and too analytic compared to the integrity of human personality, and to the perceptual openness to the world, this is not due merely to the analytic nature of computers, but rather, to the analytic nature of all modernity.

This attitude, in line with the piecemeal proposal for thought processes proposed by Newell and Simon (1972), was rather analytic. Interestingly enough, already at the height of the information processing and the production systems' symbol processing attitudes, rival, more holistic approaches were proposed on the part of the AI community. Minsky (1975) in his frame theory suggested that for computer vision – and by implication in order to understand human vision as well - one has to postulate higher order schematic packages and top-down expectations-based processing. A similar attitude was taken by Roger Schank (1975) for language material processing in the form of his conceptual dependency trees, and especially for processing human related stories in his script theory (Schank and Abelson 1977). Thus, the holistic visions, emphasizing top-down aspects in the form of the famous *frame problem* of AI, were clearly seen as an issue already at the peek of the bottom/up symbol processing times. However, they were still entirely symbolic in their processing. All of this is a modern equivalent to Boole's dream of being able to translate all intellectual activity into the language of a logical calculus. This attitude had early phenomenological criticism on the part of Dreyfus (1972), who basically claimed that machines by themselves are unable to resolve the frame problem. Real interaction, and social as well as perceptual embeddings are needed for this. Hofstadter (1979, 1987) would criticize within the AI community the traditional symbol processing attitude – this is the claim that in the final analysis all cognition is compu-

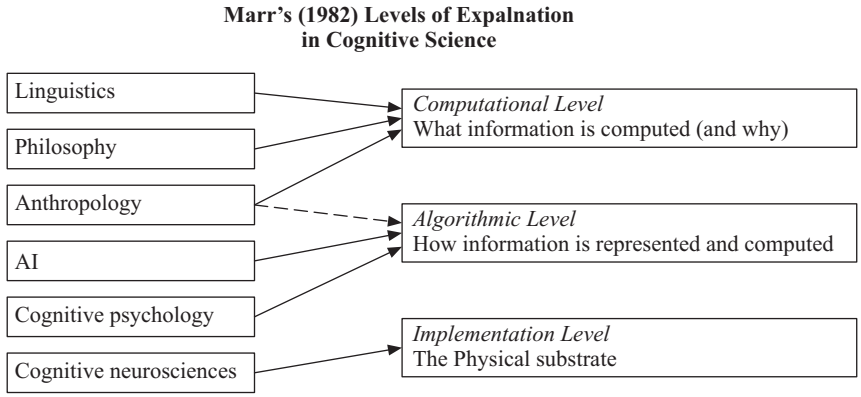


Figure 7. The three levels of explanation proposed by Marr (1982) as related to the disciplines of CogSci presented by Gentner (2010)

tation. Hofstadter also challenges the metaphorical idea made famous by Herbert Simon that whatever is interesting in cognition happens after 100 msecs, since 100 msecs is enough to recognize your grandmother. That puts perception into the shadows. Hofstadter (1979, 1987) would claim that 'subcognition' is the most interesting part of human mental life. The very fast processes of recognition, recall, and the like, and whatever happens before you recognize your grandmother, i.e., in the range below 100 msecs, is crucial to understand the mind.

As for the syntactic theory of the mind itself, it was articulated in rather fine and testable ways later on by David Marr (1982), and Noam Chomsky (1980, 1986, 2000), and framed into a detailed theory among others by Pylyshyn (1984), implying that on a theoretical level cognition is equivalent to computation. Figure 7 shows how Marr's theory that differentiated between computational, algorithmic, and implementational issues can be related even to the different constitutive disciplines of CogSci.

The new moves from the 1980s: Questioning symbol processing and the non-interpreted vision of the mind

Interpreted CogSci

One cannot be a judge of his or her own time. We have merely characterized the last two decades of CogSci to show how the internal tensions of classical CogSci have led to conceptual innovations and tensions through variation. The basic difference is ontological neutrality. The new trends in cognitive science all try to have an interpretation of cognition. They either do this interpretation in terms of neurobiology (proximal biological interpretation), or in terms of evolution (distal biological interpretation), or still, in terms of social embedding. The three approaches are summarized in Figure 8.

Once interpretation is introduced the real issue becomes reductionism versus emergency. The real novelty of present-day interpreted CogSci is twofold:

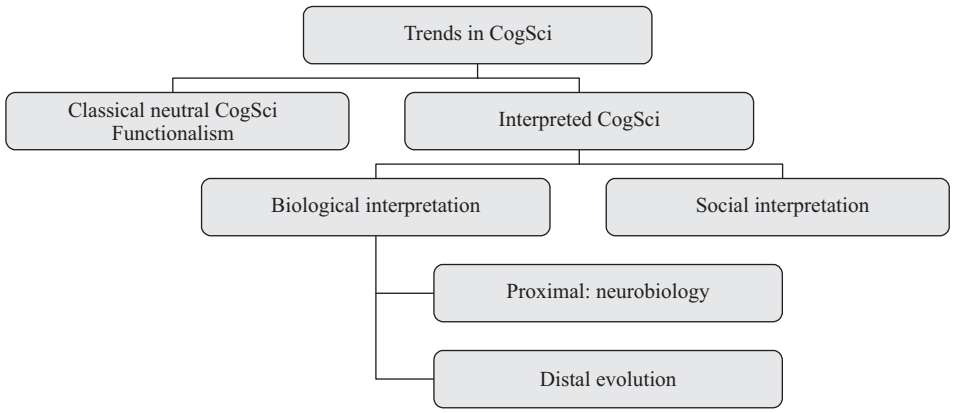


Figure 8. A different version of present-day CogSci regarding interpretation

1) The different types of interpenetration and embedding are not treated as rivals but complement each other. Think of the way mirror neurons are introduced: there is a great neuroscience discovery, followed by cognitive interpretation, which is followed by evolutionary speculations. In the same way, neuroscience interpretation of language go hand in hand with ideas about the evolution of these neural centers, and their social function.
2) All the interpretations are accompanied by a string developmental attitude. In a way, they move as if fulfilling the dreams of Piaget and Vygotsky (1978): all CogSci is basically developmental science. They propose new intricate relations between representational and non-representational cognitive systems.

Representational and non-representational approaches

Compared to traditional CogSci in the 1970s and in the early 1980s, CogSci as of today is a much more varied and divided enterprise. Table 4 shows some of these diagnostic dilemmas. We shall not go through the entire list. The 'Alternatives' column does not form a unique alternative. For instance, most of the modularists argue for symbolic processing and stick to a representationalist theory.

Rather, we pick one crucial issue, that of representations. One of the crucial dividing issues is how they interpret the role of representations in cognition. When compared to behaviorist times, everybody acknowledges today that internal factors determine our behavior in the environment. However, it has become clear that, while you admit the crucial role of the mind, you can still be a non-representationist in the sense of claiming that, instead of individuated representation, skill-like functional models are responsible for our adaptive behavior. Kästner and Walter (this volume) present a similar picture as they summarize the different skill-based, and embodied, and embedded, situational approaches to cognition, and Roy (this volume) highlights the open or hidden anti-Cartesianism of the new moves.

Figure 9 shows some alternatives regarding representationalism.

Table 4. State-of-the-art status of the cognitive enterprise: Diagnostic dilemmas (after Pléh 2009)

Traditional view	Alternatives
Unified	Modular
Symbolic	Subsymbolic
Logical, deductive	Intuitive, experiential
Sequential	Parallel
Body-independent	Body-related
Individual	Social
Static	Evolved and developing
Modellable	Inexhaustible
Truth-oriented	Directed by desires
Automatic, machine-like	Human, meaning-oriented
Knowledge-impenetrable	Knowledge-penetrable
Explicit	Implicit
Representational	Non-representational

The first great challenge to classical symbol processing cognition came from *connectionism* in the 1980s. That was a real breakthrough both technically and intellectually. It challenged the ontological neutrality by claiming a brain-based approach, and at the same time it challenged the idea of representation and symbolic processing. The *connectionist model* of information processing that has taken form from the mid-1970s on, proved to be more radical than the previous network models since it tried to reduce all representation to one

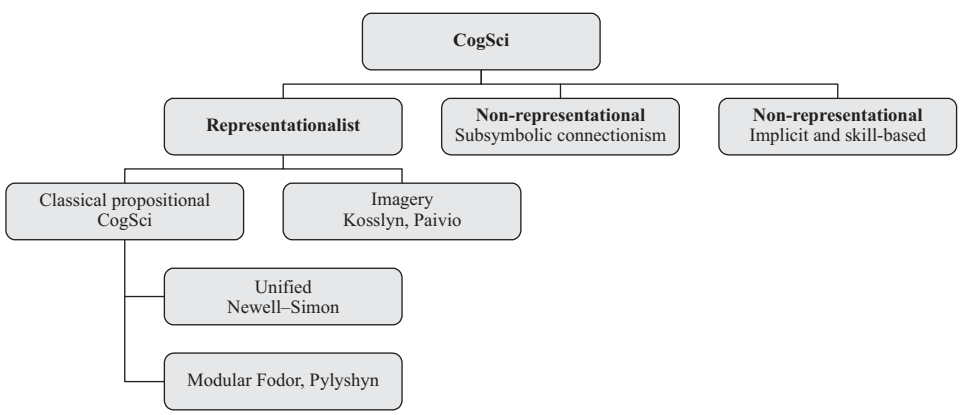


Figure 9. Versions of CogSci according to their representational stance

unitary form. In a way it kept the promise of unified cognition, but sought unification in other directions than unification based on symbol processing. As presented by Rumelhart and McClelland (1986, McClelland and Rumelhart 1986) this model for cognition had the following challenging features:

1) The model consists of *units* and their *connections.* Units are interpreted as theoretical neurons, which take the weighed sum of activation coming from their environments. Connections are positively and negatively weighted "wirings" between units.
2) *Different networks are postulated for different tasks.* One of their basic features, however, is the massive parallel and interactive organization. For instance, in the process of recognizing visually presented words, the unit corresponding to initial *T* facilitates all the units at the word level beginning with *T,* and these units, on their turn, facilitate the perception of letter *T* (its corresponding unit), and inhibit units representing other letters.
3) "*Representations* in connectionist models are patterns of activation over the units in the network" (McClelland 1988: 109). The representations are active in the sense that there is no need for a further central processor.
4) *Processing* is the unfolding of activation over time.
5) *Learning* is the modification in the strengths of connections. Both occur according to specified functions of weighing. There is no separate model for learning and for representation.
6) *Knowledge* is represented in the pattern of connections. There is nothing but connections to represent whatever knowledge we have of the world.

The last quarter of the 20th century saw interesting empirical and theoretical debates starting from the non-representationist challenge of the connectionists. Table 5 shows a classical comparison. The entire volume edited by Pinker and Mehler in 1998 clearly indicates why connectionism was such a challenge.

Table 5. The juxtaposition of connectionist and classical cognitive architecture according to Fodor and Pylyshyn (1988)

Connectionists	Classical view
Nodes	Descriptions
Only causal relations (history of excitation)	Rich relationships (language of thought)
Excitation paths	Rewriting rules
Structure-independent units (items)	Structure-dependent entities (constituents)

Fodor and Pylyshyn (1988) a generation ago clearly contrasted classical cognitivism with the connectionist challenge claiming that the basic limitation of connectionist models is their lack of structure. One can characterize this feature in many ways: "Models based on patterns of (co)excitation cannot differentiate between two concepts being active simultaneously and

them being in a given relation (like IS, PART OF etc.)". A connectionist representation has no clear syntax (lack of structure).

The associationism of connectionist models situates the human mind at the mercy of the arbitrary unsystematicity of the world: it allows any connections whatsoever (Pléh 2009).

Historians should not be judges here. It is noteworthy, however, from the point of view of the history of ideas, that after the great controversies of symbol processing versus subsymbolic networks, later synthetic work by Pinker (1991, 1997) has allowed symbolic and rule-based system for syntax, and an associative item-based network for words in language.

It seems that even approaches that are strongly committed to the overwhelming importance of structure and rules cannot ignore pattern induction and frequency-based elementaristic factors in human cognitive processes like language. Present-day models such as Ullman (2001, 2004) tend to juxtapose the two systems as a procedural system and a declarative system. In this way, the entire symbolic–subsymbolic debate becomes an issue of how to put cognition into the explicit/implicit dimension, or Gilbert Ryle's (1949) differentiation between knowing what, and knowing how, which is characterized by many even in our volume as belonging to the behaviorist past, remerges today with a central new use, similarly to the ideas of Michael Polanyi (1966) about the tacit dimension of knowledge. Classical information processing machines and the corresponding psychology would be criticized for singling out the world of knowing what while the domain of knowing how, i.e., the domain of skills would be critically important.

Figure 10 shows the how other alternatives, namely Gibson, ecological psychology, and embodiment, challenged the unifying ideas of CogSci according to Gentner (2010) (see also Kästner and Walter, as well as Roy and Reichelt and Rosmanith, and, for Gibson, Brook [this volume]. These challenges have to be taken seriously, but without the hope of ever having a

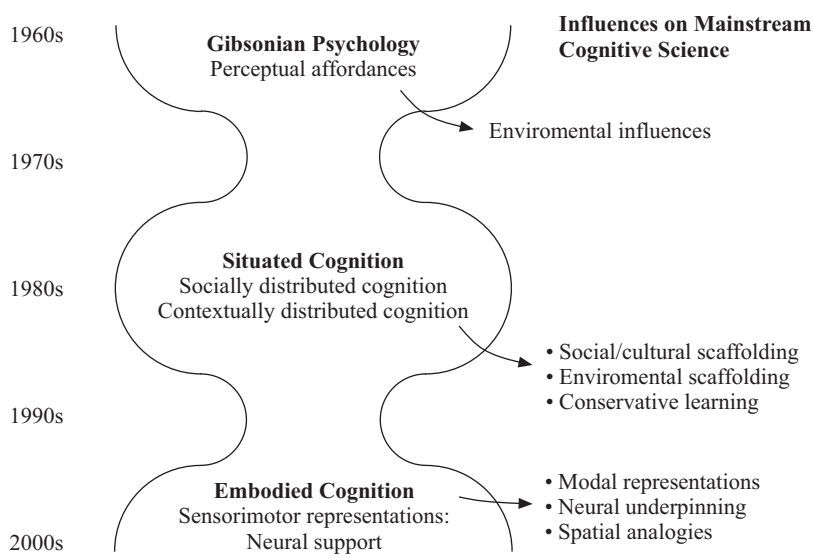


Figure 10. The different real-life challenges to classical CogSci according to Gentner (2010)

unified single model of cognition. The mind itself might be much more varied, more than our theoretical models of it.

The controversial issues as represented by some of the papers

Lena Kästner and Sven Walter's paper "Historical Perspectives on the What and Where of Cognition" represents a historically informed look at the current controversies provoked by what the authors call "the Where-question of cognition", i.e., the question whether cognition resides in the brain, in the brain and the body, or extends beyond human brain and the body. The paper reveals that the different answers to the Where-question provided by the proponents of the classical approaches to cognition on the one hand (i.e., the proponents of GOFAI and connectionism), and the proponents of the dynamical systems approach and situated cognition approach on the other hand, have resulted from taking different stances at the What-question of cognition, the question "What is cognition?" The different parties have not succeeded to agree even on whether "cognition" is a natural kind term, a cluster term, or an umbrella term. "The situation seems rather bleak", the authors conclude, but the awareness of the available options as they have been outlined in their paper "may help to guide cognitive scientific research and eventually enable us to answer both the What- and the Where-questions" of cognition.

Jean-Michel Roy's paper entitled "From Cognitive Cartesianism to Cognitive Anti-cartesianism" starts with what some may recognize as a rather severe judgment: that we still lack a historical account of cognitive science that is comparable in "levels of factual details, theoretical depth and methodological rigor" with what the professional historians of other fields of knowledge have offered so far to their public. As a step toward improving that situation, Jean-Michel Roy suggests a set of 6 hypotheses which, according to him, provide a comprehensive account of "where cognitive science research is really going". Roughly, the central tendency elicited by these hypotheses is outlined by Roy as a departure from classical cognitivism (named also "cognitive cartesianism"), which the author associates with 16 main principles. Roy insists that "every major transformation that marked the development of cognitive science since cognitivism … puts directly into question one or more of these principles". He admits at the end of the paper that the views he argues for are to be empirically tested by means of "a detailed historical investigation of the development of cognitive science" but independently of that, he says, they can be used as an "interpretative grid" for making sense and order of the chief historical facts.

Bálint Forgács's paper "The Right Hemisphere of Cognitive Science" does not directly deal with the history of cognitive science: it provides a rather unusual and at the same time highly suggestive perspective from which the history itself as well as its possible continuation could be estimated. The main idea is that the different approaches in psychology and in cognitive science could be linked metaphorically to the functions of the different regions of the brain. Thus, for example, Gestalt principles "could be linked metaphorically to the right hemisphere's posterior regions". The central ideas of early cognitive science (the computational metaphor) refer to the left hemispherical logical and sequential processes. Nevertheless, just as the processes of one area "cannot be mapped on the whole brain completely", none of the existing approaches can serve as a basis for a complete theory of cognition. The heuristic

value of this unusual brain metaphorical perspective on cognitive science is that knowing the currently popular approaches and their metaphorical brain correlates, one could predict what kind of complementary approaches are needed for a better account of the extremely complex system of cognitive processes.

In her paper "Understanding the Rational Mind: The Philosophy of Mind and Cognitive Science", Olga Markič explicates the philosophical roots of current cognitive science agendas. All of them, the author says, are trying to overcome the Cartesian divide between physical/mechanical and mental/rational behavior. The classical symbolic paradigm which was launched at the birth of cognitive science and which was organized around the primary task to reveal how mental states are physically realized did provide *in principle* a solution to Descartes' challenge but failed to implement it *in practice.* The famous frame problem is probably the clearest demonstration of the failure of the computational–representational theory to provide detailed mechanistic account of rational agency. Although successfully overcoming the traps of the frame problem, the rival bottom-up approaches are still far from being recognized as the proper tools for bridging the physical/mechanical–mental/rational gap. The paper ends with the conclusion that it is possible in principle to question Descartes' philosophical settings which underlie current cognitive science agenda but "that will lead us to another story", a story that will obviously transcend the real story of "the mind's new science".

An alternative frame: The importance of the European heritage

In this section, we intend to survey in a rather cursory way some of the European, mainly continental traditions during the last 150 years that have reemerged in present-day cognitive science as a hidden past. The survey is impressionistic since it mainly starts from the consideration of the contemporary scene. It is not a real historical analysis, rather, it starts from the framework of a history of ideas. Brook (this volume) takes a similar attitude when he outlines the prehistory from Descartes, through Locke and Kant, to Frege and Freud. We look for ancestry and traditions, rather than analyzing the actual historical scenes in depth. The

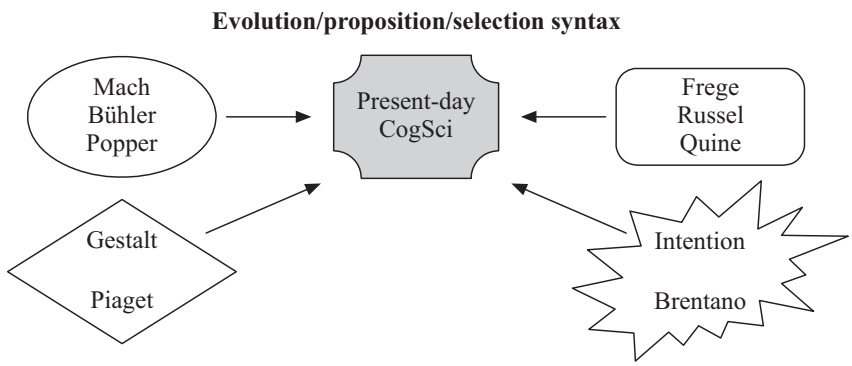


Figure 11. The forgotten or neglected European traditions of CogSci

intention is to show that some of the hidden dimensions of Classicism go back to traditional, mainly philosophical issues, and that the emerging new attitude also has a historical ancestry. Figure 11 shows the four European traditions that are the important immediate ancestors of some present-day concerns.

A forgotten evolutionary positivist forerunner: Ernst Mach and how he reappears in evolutionary interpretations of cognition

The presence of Mach is hardly felt in the present-day cognitive scene, while he was responsible both for the positivist image of man so crucial for AI, and for the evolutionary interpretation of cognition. This neglect is true regarding the philosophical interpretations of evolutionary theory as well.

Leszek Kolakowski (1968: 155), in his treatment of positivist philosophy gave a summary of Mach's views that is relevant for the history of CogSci as well:

> [W]e are especially struck by the following features [in Mach]:
> 1) the philosophical destruction of the subject;
> 2) the biological and practical conception of cognitive functions, reduction of intellectual behavior to purely organic needs, and renunciation of 'truth' in the transcendental sense;
> 3) desire to get back to the most primitive concrete datum, to a 'natural' view of the world not mediated by metaphysical fictions.

Let us interpret these three issues highlighted by Kolakowski from Mach's point of view – Mach being a not too distant forerunner of contemporary CogSci.

1) Ernst Mach initiated a remarkable tradition of claiming *continuity between everyday cognition and science.* As a further step along the line of this unity, epistemology, and the psychology of cognition, disregarding the warning of Kant, belong together according to Mach. Lana (1976) is one of the exceptional present-day historians of psychology who noticed the importance of this fact. This feature is relevant for our purposes here since this continuity between epistemology and psychology is also true for almost the whole present-day cognitivism: there is no dividing line between science and ordinary cognition, and similarly, there is no dividing line between epistemology and the empirical sciences studying human cognition. This feature, which was put into a general theory in the *naturalized epistemology* of Quine (1969), is a defining feature of classical CogSci. That is why psychology and philosophy collapse in a way into CogSci.
2) *Cognition is interpreted by Mach in an evolutionary framework.* That is true for the approach taken today by Dennett (1994), Pinker (1997), Tooby and Cosmides (1992), and by Tomasello (1999, 2009). Mach saw continuity between the neural and evolutionary interpretation of cognition, interpreting the former in the framework of the latter: "In several of his technical writings – including the *Contributions to the analysis of sensations,* he presented an evolutionary image of several aspects of sensory representation, including colors and dimensions of space perception. The logical structure of his argumentation is rather interesting even for present-day cognitive studies. In accordance with his general

monistic commitments, he believes in a total psycho-physiological parallelism: for each aspect of sensory experience, there is a corresponding physiological structure. That structure has to be explained by evolutionary considerations, on its turn, as an adaptation to the environment. Thus, in principle, Mach claims a dual biological anchorage for the mental: not only a short-range physiological, but also a long-range evolutionary account as well" (Pléh 2008: 25).

3) *The disintegration of the substantial Ego concept.* In this regard, Mach is a positivist, who at the same time treats the self-related concepts as soft notions, not unlike the stances promulgated by Dennett (1987) a century later. The positivist credo is clear:

> The antithesis of Ego and world, sensation (phenomenon) and thing then vanishes, and we have simply to deal with the connexion of the elements… The primary fact is not the I, the Ego, but the elements (sensations). The elements constitute the I. ... when I die... only an ideal mental–economical unity, not a real unity, has ceased to exist. (Mach 1897: 11, 19–20)

This positivist credo is accompanied by a functionalist evolutionary anchorage. The Ego is not a mere illusion, but something derived and postulated, reinforced for practical purposes by the principle of economy which is basically a Darwinian idea in Mach:

> We are prepared, thus, to regard ourselves and every one of our ideas as a product and a subject of universal evolution… We feel that the real pearls of life lie in the ever changing contents of consciousness, and that the person is merely an indifferent symbolic thread on which they are strung. (Mach 1910: 234–235)

This soft concept of Ego with which we have to live, without taking it to be an ontological starting point, has two kinds of anchorages for Mach. The first comes from our general tendency to abbreviate repeated complexes of sensations with names, which appears in our notion of physical bodies as well: "That which is perceptually represented in a single image receives a *single* designation, a *single* name" (Mach 1897: 3).

The other anchoring point is found by Mach in the concept of body: "As relatively permanent, [it] is exhibited, further, that complex of memories, moods, and feelings, [are] joined to a particular body (the human body), which is denominated the "I" or "Ego" (Mach 1897: 3).

This early and, one could say, premature evolutionary vision of cognition presented by Mach has two lines that lead to the work of the present-day evolutionary cognitivist philosopher Daniel Dennett.

1) The first is the whole notion of carefully *continuing the Darwinian path* and studying behavior patterns, and stimulus preferences as the results of evolution. The Viennese tradition of evolutionary thought in the work of Konrad Lorenz (1941, 1965), himself a student of Bühler, through the elaboration of comparative ethology, transmitted this idea up to the present-day concern with evolutionary explanations for more inner aspects of cognition. See details below.

2) Mach introduced the general idea of *selectionist epistemology:* hypotheses and trials characterize not only science, but all our everyday cognition. We shall see below how this was taken up by Bühler, and then by Popper.

Interestingly enough, the views of Mach – *ceteris paribus,* of course there are 150 years of empirical and conceptual research between them – emerge as a key model for the 'New Synthesis' proposed by Dennett (1987, 1991, 1994) in his radical evolutionary vision of CogSci (see Pléh 2009). Some of the basic ideas entertained by Dennett that concern us in this respect might be summarized as follows:

1) *Some of our cognitive achievements are related to attitudes,* to relatively soft stances towards the world like the intentional stance (Dennett 1987) that are 'soft' in the sense that they can be given an instrumental interpretation.
2) We are entitled to use them because *they work pragmatically.*
3) This pragmatic feasibility is related to the fact that *they were formed in our evolutionary history* (Dennett 1991, 1994).
4) There are *no mysterious or essentialistic static points,* or preset goals neither in the structure and the processing of the mind, nor in its evolution.
5) Self and consciousness are neither fixed starting points nor identifiable inner places. Rather, *self-related notions are shifting abstract entities* like gravitational centers which are useful for our orientation (Dennett 1991; Dennett, and Kinsbourne 1992).

The history of the evolutionary tradition in CogSci: Evolutionary CogSci traditions and the Vienna school

In this section, we attempt to survey the also hidden tradition that continued the evolutionary approach to cognition proposed by Mach as part of a broader general vision of behavioral evolution and selectionist thought. The Darwin bicentennial raised awareness again to the fact that within philosophy, biology, and psychology, there were several traditions of evolutionary thinking as related to cognition. The basic implications were already clear at the end of 19th century:

1) Mental has to have a biological function.
2) All mental should be looked in its process, in time and in unfolding.

Table 6. Psychological implications of the structure of Darwinian theory
(based on Lewontin 1970 and others; after Pléh 2008)

Darwinian principles	Psychological Darwinism
Fitness and its variability	Adaptive view of the mind: Mach, James, Dewey
Phenotypic variability	Study of individual differences: Galton paradigm
Inheritance	Selection of habits and mental traits

3) Every mental capacity or process has to be looked at from two developmental perspectives: that of animals and children.
4) Varieties and variations are key features of mental life

This affinity is indicated in Table 6.

Already in late 19th century, three types of Darwinian psychology were born out of these considerations. The most well-known is the 'hairy comparative' or 'animal psychology' that studies the phylogenesis of the human mind with different attitudes (everything is always there, there are quantitative differences among the species, or one can postulate qualitative differences among the species) already outlined by Conwy Lloyd Morgan (1894). This was and still is accompanied by the study of individual differences. However, the most interesting or surprising for the history of CogSci is the (C) variety of Darwinian psychology: relating evolutionary theory to selectionist models of epistemology.

A) Comparative: From Romanes through Thorndike and ethology to Premack et al.
B) Study of mental individual differences: From Galton to Cloninger.
C) Epistemological:
   – Armchair and experiments: Mach to Dennett.
   – Babies as knowers: Baldwin (1894, 1896) to Tomasello (1999, 2009) and Gergely (2001, Csibra and Gergely 1998).

The basic approach continued from the evolutionary tradition of Mach is the idea of separating two cycles in any changing system: the cycle of generating solution proposals, and the choice, i.e., the selection between them. In its most abstract form, this is summarized by Karl Popper (1972: 243–245) as a cyclic and gradual view of any change, the separation of idea or solution generation, on the one hand, and selection, on the other.

This general image was first outlined in Vienna by Karl Bühler, a teacher of Popper. From the point of view of the history of ideas Bühler in the 1920s and 1930s tried to overcome in a sometimes eclectic, but certainly liberal way the controversies among the internalist, the behavioral, and the culturalist approaches to cognition that was for him psychology (Bühler 1922, 1927, 1934, 1936, 1990). He belonged to the class of those Central European scholars who were looking for a meaningful unity in their science, while being aware of the divisive naturalistic and spiritualistic trends. The much cited quote below shows how relevant this attitude is even for the present divisions of the study of cognition.

> When someone raises a new topic, why does he have to look down scientifically on his neighbor? In the large house of psychology there is room for everyone; one could direct his spectacles on the skyline of values from the attic, others could at least claim for themselves the basement of psychophysics, while the walls are intended to out the entire enterprise into the causal chain of events. (Bühler 1927: 142.)

The actual substance of Bühler's proposals were founded in his *sign theoretic approach to psychology,* and in his thorough knowledge both of contemporary *Denkpsychologie* and linguistics, one the one hand, and early ethology, on the other hand. The features of his rich oeuvre can be summarized as basis theses. The evolutionary aspects are highlighted with letter type.

1) All behavior is regulated by signs. *There is no meaningless behavior.*
2) Human behavior is oriented to supraindividual meanings. *All human behavior has three aspects:* ***experience, behavior, and reference.***
3) All behavior is characterized by holistic organization aimed at species-specific signals. *Structure, meaning, and goals characterize all behaviors.*
4) An evolutionary organization of behavior is postulated with superposed levels of selection. *No barrier between the amoeba and the poet.*

Bühler, when trying to interpret the Darwinian message to psychology, postulates a universal metatheory of selection, with three levels of selection. The selectionist metatheory came as a way to integrate the different attitudes. As part of the extension of selectionist explanations to different domains, Bühler (1922) extended Mach's (1905) idea of seeing hypothesis and trial everywhere. Bühler proposed a continuity between instinct, trial and error learning, and intellect. Through the mediation of Popper, these correspond roughly to what Dennett (1994, 1996) calls Darwinian, Skinnerian, and Popperian creatures.

> For me, in Darwinism the concept of play field seems to be productive. Darwin has basically known only one such play field, while I point to three of them. […] These three play fields are: instinct, habit and intellect. (Bühler 1922: VIII.)

In this framework, animal behavior is assumed to be intentional and purposeful. Intentions and signs organize animal behavior as well as human mental life: there is no demarcation line between human mentality and animal mental life. Intention-based, teleological, and holistic organization is true of all behaviors, and it creates unity between the work of biology and that of the psychology. Table 7 shows how the different levels of behavioral selection were distinguished by Bühler.

Table 7. Three levels and pools of selection according to Bühler (after Pléh 2009)

Features	Instinct	Habit	Intellect
Pool of selection	Individuals	Behaviors	Thoughts
Roads to selection	Darwinian selection	Reinforcement	Insight
Proofs	Species-specific behavior	Associations, new combinations	Detour
Representative author	Volkelt, Driesch	Thorndike	Köhler
Organization	"Naturplan"	Associative net	Mental order

Karl Popper, when he continued this Darwinian heritage of Karl Bühler, clearly spelled out the message: in his Spencer lectures, Popper characterizes the analogy of thinking with natural selection in the following way: "the growth of our knowledge is the result of a process closely resembling what Darwin called 'natural selection'; that is *the natural selection of hypotheses*" (Popper 1972: 261).

Table 8. The eleven different types of selection moving upwards from Darwinian genetics to science (Campbell 1974)

Domain	Example
Science	Hypothesis–Solution–Choice
Cultural accumulation	Selection in technology
Language	Language variation
Observation and imitation	Social insects
Thought supported by memory	Imagery-based solutions
Visually supported thought	Köhler: insights in apes
Habit	Rearranging control systems
Instinct	Organismic perceptual systems
Vicariating locomotion	Echolocation
Problem-solving without memory	Tropisms
Genetic adaptation	Genetic variation and change

Later on, during the 1950s, the *evolutionary epistemology* as it came to be cited, has worked out the selectionist vision in further details. Donald Campbell (1974), the excellent methodologist and social psychologist, as a follower of Popper gave the most exhaustive list of what was available before the onset of modern cognitive science as levels of behavior and levels of selection. This is shown in Table 8.

Bühler (1922) and Popper (1972), as well as Campbell (1974) for instance see an emergent evolutionary relationship between these levels: they are not only homologues, but they are also assumed to have a common causal history. Popper and Campbell are mainly forgotten with their evolutionary message in present-day self representations of cognitive science. That is why we can refer to the entire tradition as the forgotten Viennese tradition. For instance, the relatively new MIT cognitive science encyclopedia has only four references to Popper in a volume of a thousand pages. Popper figures in the entries of *consciousness* and *emergentism,* as well as regarding induction, and his methodological critic of psychoanalysis. However, his evolutionary epistemology as a general approach figures nowhere (Wilson, and Keil 1999).

This is compensated by the philosophical tradition. Through the mediation of Popper, similar ideas show up in later extended Darwinian approaches as well. Bühler's three levels roughly correspond to what Dennett (1994, 1996) calls the tower of selection: Darwinian, Skinnerian, and Popperian creatures. The idea of a selectionist approach to cognition reemerges within CogSci during the last two decades of interpreted CogSci.

The rebirth of intentionality in modern cognitivism

Present-day cognitive theory has also observed the rebirth of another party Viennese and certainly European heritage: that of *intentionality.* It is a triviality from classical rhetoric, from the time of Quintilianus on that the essence of signs is that they stand for something or someone. The scholastic theory of intentionality refreshed in the late 19th century in the work of Brentano started to put this relationship not only in focus with the theory of signs but with a general theory of *representational intentionality.* As Brentano (1874) claimed, intentionality, 'aboutnesss' is the defining feature of the mental. Mental events do not stand alone: they always indicate something beyond themselves.

For Brentano himself, intentionality was a conceptual issue. While his contemporaries such as Wundt saw the defining feature of the mental in its accessibility for introspection, for Brentano the essential aspect was the 'aboutness', the fact that mental phenomena never stand in themselves, they always relate to something else.

> Every mental phenomenon is characterized by what the Scholastics of the Middle Ages called the intentional (or mental) inexistence of an object, and what we might call, though not wholly unambiguously, reference to a content, direction toward an object, … or immanent objectivity. Every mental phenomenon includes something as object within itself... (Brentano 1874: 88)

> There is no hearing without something heard, belief without something believed, hope without something hoped, and a striving without goal. (Brentano 1874: 88)

> This holds true for emotions as well:

> One is happy *because of something,* … and we say I am happy *for this, this troubles me*... (Brentano 1874: 89)

> Three types of intentional relation were differentiated by Brentano:

1) *Ideation* (in sensation and imagining).
2) *Judgment* (existence–non existence, truth).
3) *Like–hate.*

However, for Brentano – and that is why he becomes so crucial for the representational theories of the mind – ideation is the basis for cognition:

> The representational process is the basis not only of judgment, but of desire and all other mental acts. Nothing can be judged, or desired, or feared without being represented. (Brentano 1874: 89)

Modern cognitive science took up Brentano's path of a century later, in the 1970s. For them, the issue of intentionality always relates to the issue of the *reducibility of mental qualities,* and thus separates even present-day cognitivists into three groups:

1) *Intentional realists:* Intentionality is an irreducible quality thus providing for the boundary conditions of naturalism (Searle 1980, 1992).
2) *Radical behaviorists:* Intentionality as such does not exist.
3) *Intentional instrumentalists:* Intentionality does exist but it can be given a naturalistic reduction (Dennett 1987).

For the realists, the essence of the representational theory is the proposal that in thinking we always have to do with relations between representations, i.e., with representational systems. Representations in simple cases refer to something in the world, but at the same time have a peculiar relationship to each other. From a given representation, the validity of other representations follows. In representational systems – to allude to the famous argumentation of Jerry Fodor (1990) –, causal and implicational relations both hold. From the truth of the sentence *Frankie forgot that Mary closed the door* logically follows the truth of the sentence *Mary closed the door.* This is, however, not a causal consequence, but is derived from the structural and meaning relations between the sentences involved. Similarly, if on a figure, C is above B, and B is above A, it follows that C is above A. This is not some kind of causal relationship between the propositions about the figure, but comes from the structure of the images themselves (Fodor 1990).

In the rebirth of the interest towards intentionality in present-day CogSci, several novelties can be observed. They all relate to the issue of the boundary conditions of intentionality (do our computers or dogs think?), and to the entire naturalization of cognition. Rather then taking intentionality as a categorial starting point, present-day cognitivists attempt to postulate levels of intentional organization in an evolutionary context.

0) Teleonomy in biology (reflexes, tropisms) provided an evolutionary explanation.
1) Behavioral equivalence relations in animal life.
2) Attribution of goals and intentions both to others and to objects. See Gergely, Nádasdy, Csibra, and Bíró (1995) on intentional stance in infants.
3) Separation of representation and reality: attributing belief in the theory of mind.
4) Free and flexible uses of the intentional stance as a stance.

The Frege tradition

During the late 19th century, in the heydays of German academic dominance, two attitudes were outlined towards the study of thinking. Wilhelm Wundt outlined a sensualistic theory of thought hoping that thought can be assigned to individual minds, and it could be given a sensation- plus association-based interpretation. Gottlob Frege (1884, 1892) on the other hand, claimed that thought cannot be assigned to individual minds, and it cannot be analyzed with images and associations. Rather, it has to be interpreted in a Platonistic supraindividual way, and it has a propositional structure.

This has lead to many fierce debates within German academia about ‘psychologism’ (Kush 1995), and this basically separated psychology from the logical enterprise. In the latter, through the mediation of Husserl (1900) and Russell, Frege has become a constant star over the 20th century. In a peculiar way, CogSci changed this state of affairs. With the advent

of the representational theory of mind (RET), Frege had a victorious return to psychology. Frege was right in characterizing thinking in propositional terms. He was wrong, however, in assuming that individual thinking can only be given an associative interpretation. Rather, the individual knower is assumed to follow the propositional calculus, and the individual mind is a host to propositions.

Most psychologists amongsts his contemporaries stayed with the sensory–associative model. There was a minority in the 1990s, however, who took up the challenge of Frege in the Würzburg school, out of which Karl Bühler also came. According to them, individual mental processes point towards supraindividual mental organizations. One can touch upon this even in the laboratory by emphasizing the non-sensual ('unanschauliches') character of thoughts emphasized especially by Bühler (1908). Our thoughts, as the leader of the Würzburg school emphasizing a logical approach to psychology pointed out, are always directed towards something beyond itself, something objective, to thought structures (Külpe 1912).

> Thought thus can be directed to objects that are of a very different nature from it and become by the fact of being representation pure mental contents, or pure thoughts. Experimental research not only clarified this to be the case[3], but at the same time it showed that represented objects can have different status […], and therefore their relations to thinking might be of a different kind. This should be understood here as a differentiation between concepts and objects, and among these latter ones ideal, real and represented objects. (Külpe 1912: 1088)

This was, however, a weak minority at the time. The time of Frege came in the 1970s to psychology, in the gown of CogSci. At a time when we constantly emphasize skills and images, we should not forget that, when it was first proposed, this propositional approach in the works of Frege (1892) and Husserl (1901–1902) was clearly opposed to a traditional sensualistic vision. One could even claim that the great turn-of-the-century revolution in the philosophy of mind (i.e., Frege and Husserl) with their Platonic propositions made possible the use of the same constructions for machines and for the human mind as well – one hundred years later. This could be called Freges's revenge: propositions are the first weapons of an anti-psychological campaign, to become reintegrated into psychology almost a century later (Pléh 2009).

In this regard, modern theories of cognition can be classified according to the representational relations that the different schools entertain. Some claim, of course, that there are no representational relations whatsoever. Some of them can be labeled mean behaviorists, some of them, however, more sophisticated non-representational theoreticians, who emphasize skills and embodiment. There are others according to whom representational relations are principally of a linguistic nature. This is referred to today as the syntactic theory of mind. Human thinking is interpreted as a merely formal system. In a syntactic engine, sentences lead to other sentences, and in our mind, propositions lead to other propositions. Ideas themselves would be organized in a sentence-like manner (Fodor 1975). This is a propositional theory of the human mind, which could aptly be called, following the function-like proposal about logical structure advanced by Gottlob Frege, as the Frege-model of human thought (Frege 1984).

[3] I.e., in the interpretation of the Würzburg group.

In the early 20th century, followers of the classical sensual ideas claimed that the basic vehicle of human thought would be *images.* Thinking is always image-like, and the sensual content of images carries meaning. The two theories are still with us as two rival approaches of representational theories had characteristic debates in several areas. The propositional camp is represented by Pylyshyn (1984). Some people would question whether representations would indeed always be of a propositional form (Kosslyn 1980, 1994).

The interesting hidden European continental tradition is that, in some regards, the developments foreshadowed the conceptual heritage of intentionality, the logical organization of representations, and the issue of their origins. These are the central issues especially in their interrelationship possibility of present-day cognitive science as well.

References

Arbib, M., Érdi, P., and Szentágothai, J. (1997) Neural Organization: Structure, Function, Dynamics. Cambridge (MA): MIT Press.

Ashby, R. (1956) *An Introduction to Cybernetics*. London: Chapman, and Hall.

Baldwin, J. M. (1894) Mental Development in the Child and the Race: Methods and Processes. New York: Macmillan.

Baldwin, J. M. (1896) A new factor in evolution. American Naturalist 30, 441–451, 536–553.

Barsalou, L. W. (2010) Introduction to the 30th anniversary perspectives in cognitive Science: Past, present, and future. Topics in Cognitive Science 2, 322–327.

Block, N. (1980–1981) Readings in the Philosophy of Psychology. Vols. I, and II. Cambridge: Harvard University Press.

Boden, M. (2006) *Mind as Machine: A History of Cognitive Science*. Vols. I, and II. Oxford: Oxford University Press.

Brentano, F. (1874) Psychologie vom empirischen Standpunkt. Leipzig: Meiner.

Broadbent, D. (1958) Perception and Communication. London: Pergamon.

Bruner, J. S. (1973) Beyond the Information Given. New York: Norton.

Bruner, J. S. (1983) *In Search of Mind: Essays in Autobiography*. New York: Harper, and Row. Online version: http://www.questia.com/library/book/in-search-of-mind-essays-in-autobiography-by-jerome-bruner.jsp.

Bruner, J. S. (1997) Will cognitive revolutions ever stop? In: D. M. Johnson, and C. E. Erneling (ed.) The Future of the Cognitive Revolution. New York: Oxford University Press, 279–292.

Bruner, J. S., Goodnow, J., and Austin, G. (1956) A Study of Thinking. New York: Wiley.

Bruner, J. S., Oliver, R. R., and Greenfield, P. M. (1966) Studies in Cognitive Growth. New York: Wiley.

Bühler, K. (1908) Tatsachen und Probleme zu einer Psychologie der Denkvorgänge. II. Über Gedankenzusammenhänge. II. Über Gedankeneinnerungen: Archiv für die gesammte Psychologie 12, 1–23, 24–92.

Bühler, K. (1922). Die geistige Entwicklung des Kindes. Jena: Fischer. 3rd edition.

Bühler, K. (1927). *Die Krise der Psychologie.* Leipzig: Bart.

Bühler, K. (1934) *Spractheorie.* Jena: Fischer. New English translation: (1990) *Theory of Language: The Representational Function of Language.* Translated by D. F. Goodwin. Amsterdam: John Benjamins Publishing Company.

Bühler, K. (1936) Die Zukunft der Psychologie und die Schule. Wien, and Leipzig.

Campbell, D. T. (1974) Evolutionary epistemology. In: P. A. Schlipp (ed.) *The Philosophy of Karl Popper.* La Salle, and Open Court, 413–463.
Chomsky, N. (1980) *Rules and Representations.* New York: Columbia University Press.
Chomsky, N. (1986) Knowledge of Language. Its Origins, Knowledge, and Use. New York: Praeger.
Chomsky, N. (2000) *New Horizons in the Study of Language and Thought.* Cambridge: Cambridge University Press.
Cicourel, A. (1974) Cognitive Sociology: Language and Meaning in Social Interaction. New York: Free Press.
Colby, B., Fernandez, J. W., Kronenfeld, D. B. (1981). Toward a Convergence of Cognitive and Symbolic Anthropology. New York: Blackwell Publishing.
Csányi, V. (1988) Evolutionary Systems and Society: A General Theory. Durham: Duke University Press.
Csibra G., and Gergely, G. (1998). The teleological origins of mentalistic action explanations: A developmental hypothesis. *Developmental Science* 1, 255–259.
D'Andrade, R. (1995). *The Development of Cognitive Anthropology.* Cambridge: Cambridge University Press.
Danziger, K. (1990) *Constructing the Subject.* New York: Cambridge University Press.
Dennett, D. (1987) *The Intentional Stance.* Cambridge (MA): MIT Press.
Dennett, D. (1991) *Consciousness Explained.* Boston: Little Brown.
Dennett, D. (1994) *Darwin's Dangerous Idea.* New York: Simon, and Schuster.
Dennett, D., and Kinsbourne, M. (1992) Time and the observer: The where and when of consciousness in the brain. *Behavioral and Brain Sciences* 15, 183–247.
Dougherty, J. (1985, ed.) *Directions in Cognitive Anthropology.* Illinois: University of Illinois Press.
Dreyfus, H. (1972) *What Computers Can't Do: A Critique of Artificial Intelligence.* New York: Harper. 2nd ed.: (1979) San Francisco: Freeman.
Dupuy, J. P. (1994) *Aux origines des sciences cognitives.* Paris: Dandcouvertes.
Dupuy, J. P. (2009): On the Origins of Cognitive Science: The Mechanization of the Mind. Cambridge (MA): MIT Press.
Fodor, J. (1985): Fodor's guide to mental representation: The intelligent auntie's vademecum. *Mind*, *94*, 76-100.
Fodor, J. (1990): *A theory of content and other essays.* Cambridge, Ma.: MIT Press
Fodor, J. (1975) *The Language of Thought.* Cambridge (MA): Harvard University Press.
Fodor, J. A., and Pylyshyn, Z. W. (1988) Connectionism and cognitive architecture. *Cognition* 28, 3–71.
Frege, G. (1884/1950) *The Foundations of Arithmetic*. Translated by J. L. Austin. New York: Philosophical Library.
Frege, G. (1892/1984) *Collected Papers on Mathematics, Logic and Philosophy.* Oxford: Oxford University Press.
Gadamer, H.-G. (1983) Citizens of two worlds. In: H. G. Gadamer. *On Education, Poetry, and History: Applied Hermeneutics.* Albany (NY): State University of New York Press.
Gardner, H. (1985) The Mind's New Science: A History of the Cognitive Revolution. New York: Basic Books.
Gentner, D. (2010) Psychology in cognitive science: 1978–2038. *Topics in Cognitive Science* 2, 328–344.
Gergely, G. (2001). The development of understanding self and agency. In: U. Goshwami (ed.) *Handbook of Childhood Cognitive Development.* Oxford: Blackwell.
Gergely, G., Nádasdy, Z., Csibra, G., and Bíró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition* 56, 165–193.

Goldman, A. J. (1993, ed.) *Readings in Philosophy and Cognitive Science.* Cambridge (MA): MIT Press.
Haggbloom, S. J. et al. (2002) The 100 most eminent psychologists of the 20th century. *Review of General Psychology* 6 (2), 139–152.
Harnad, S. (1978) Editorial. Behavioral and brain sciences. http://www.ecs.soton.ac.uk/~harnad/Temp/Kata/bbs.editorial.html.
Harnish, R. M. (2002) Mind, Brains, Computers. An Historical Introduction to the Foundations of Cognitive Science. Malden (MA): Blackwell.
Harré, R. (2002) Cognitive Science. A Philosophical Introduction. London: Sage.
Hebb, D.O. (1948): *The organization of behavior*. New York: Wiley
Hofstadter, D. R. (1979) Gödel, Escher, Bach: An Eternal Golden Braid. New York: Basic Books.
Hofstadter, D. R. (1987) Cognition, subcognition. Sortir du rêve de Boole. *Le Débat* 47, 26–44.
Husserl, E. (1900–1901) *Logische Untersuchungen.* Vols. I, and II. Halle: Fisher.
Jackendoff, R. (1992) Languages of the Mind: Essays on Mental Representation. Cambridge (MA): MIT Press.
Jackendoff, R. (2009) *Language, Consciousness, Culture.* Cambridge (MA): MIT Press.
Johnson, D. M., and Erneling, C. E. (1997, eds.) *The Future of the Cognitive Revolution.* New York: Oxford University Press.
Kolakowski, L. (1968) Positivist Philosophy: From Hume to the Vienna Circle. New York: Doubleday.
Kosslyn, S. M. (1980) *Image and the Mind.* Cambridge (MA): Harvard University Press.
Kosslyn, S. M. (1994) *The Resolution of the Imagery Debate.* Cambridge (MA): Harvard University Press.
Kuhn, T. (1970) *The Structure of Scientific Revolutions.* 2nd, enlarged ed. Chicago: University of Chicago Press.
Külpe, O. (1912) Über die moderne Psychologie des Denkens. *Monatschrift für Wissenschaft, Kunst und Technick* 14, 1070–1100.
Kusch, M. (1995) Psychologism: A Case Study in the Sociology of Philosophical Knowledge. London: Routledge.
Lachman, R., Lachman, J. L., and Butterfield, E. C. (1979) *Cognitive Psychology and Information Processing: An Introduction.* Hillsdale: Erlbaum.
Lana, R. E. (1976). *The Foundations of Psychological Theory.* Hillsdale (NJ): Lawrence Erlbaum.
Lewontin, R. C. (1970) The units of selection. *Annual Review of Ecology and Systematics* 1, 1–18.
Lorenz, K. (1941) Kant's Lehre vom apriorischen in Lichte gegenwartiger Biologie. *Blatter für Deutsche Philosophie* 15, 94–125.
Lorenz, K. (1965) *Evolution and Modification of Behavior.* Chicago: University of Chicago Press.
Mach, E. (1897). *Contributions to the Analysis of Sensations.* Translated by C. M. Williams. Chicago (IL): Open Court.
Mach, E. (1905/1926) *Knowledge and Error. Sketches on the Psychology of Enquiry.* Dordrecht: D. Reidel (1976). Original: *Erkenntniss und Irrtum.* Leipzig: Bart, 1905. 5th edition: 1926. MacMillan Company.
Mach, E. (1910) *Popular Scientific Lectures.* 4th edition. Translated by Thomas J. McCormack. Chicago (IL): Open Court.
Marr, D. (1982) *Vision.* San Francisco: Freeman.
McClelland, J. L. (1988) Connectionist models and psychological evidence. *Journal of Memory and Language* 27, 107–123.
McClelland, J. L., and Rumelhart, D. E. (1986) *Parallel Distributed Processing.* Vol. 2. Cambridge (MA): MIT Press.

McCulloch, W. S., and Pitts, V. (1943) A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics* 5, 115–133.
Mehler, J., and Franck, S. (1995, eds.) *Cognition on Cognition.* Cambridge (MA): MIT Press.
Miller, G. A. (1962) Psychology: The Science of Mental Life. New York: Harper.
Miller, G. A. (2003) The cognitive revolution: A historical perspective. *Trends in Cognitive Sciences* 7, 141–144.
Miller, G. A., Galanter, E., and Pribram, K. H. (1960) *Plans and the Structure of Behavior.* New York: Holt.
Minsky, M. (1975) A framework for representing knowledge. In: P. Winston (ed.) *The Psychology of Computer Vision.* New York: McGraw–Hill.
Minsky, M., and Papert, S. (1969) *Perceptrons.* Cambridge (MA): MIT Press.
Morgan, C. L. (1894) Introduction to Comparative Psychology. London: Scott.
Newell, A. (1980) Physical symbol systems. *Cognitive Science* 4, 251–283.
Newell, A. (1989) *Unified Theories of Cognition.* Cambridge: Harvard University Press.
Newell, A., and Simon, H. (1972) *Human Problem Solving.* Englewood Cliffs (NJ): Prentice-Hall.
Newell, A., Rosenbloom, P. S., and Laird, J. E. (1989) Symbolic architectures for cognition. In: M. I. Posner (ed.), *Foundations of Cognitive Science*. Cambridge (MA): MIT Press, 93–131.
Norman, D. (1981, ed.) *Perspectives in Cognitive Science.* Hillsdale: Erlbaum.
Osherson, N. et al. (1990, ed.) *An Invitation to Cognitive Science.* Cambridge (MA): MIT Press.
Piaget, J. (1963) L'explication en psychologie et le parallélisme psycho-physiologique. In: P. Fraisse, and J. Piaget (eds.) *Traité de psychologie expérimentale. Fascicule I. Histoire et méthode.* Paris: PUF, 137–184.
Piaget, J. (1970). Epistémologie des sciences de l'homme. Paris: Gallimard.
Pinker, S. (1991) Rules of language. *Science* 253, 530–555.
Pinker, S. (1997) *How the Mind Works.* New York: Norton.
Pinker, S., and Mehler, J. (1988, ed.) *Connections and symbols.* Cambridge (MA): MIT Press.
Pléh, Cs. (1998) An Introduction to Cognitive Science. Budapest: Typotex.
Pléh, Cs. (2008) *History and Theories of the Mind.* Budapest: Akadémiai Kiadó.
Pléh, Cs. (2009) Darwin and the nature–culture continuity issue regarding culture. Invited Talk at the ESF–COST Research Conference Complex Systems and Changes. *Darwin and Evolution: Nature–Culture Interfaces.* Sant Feliu de Guixols, Spain, 15–20 September, 2009.
Polanyi, M. (1966) *The Tacit Dimension.* London: Routledge.
Popper, K. R. (1972) *Objective Knowledge: An Evolutionary Approach.* Oxford: Clarendon Press.
Posner, M. (1989, ed.) *Foundations of Cognitive Science.* Cambridge (MA): MIT Press.
Putnam, H. (1960) Minds and machines. In: S. Hook (ed.) *Dimensions of Mind.* New York: New York University Press, 148–180. Reprinted in Putnam (1975).
Putnam, H. (1975) *Mind, Language and Reality.* Cambridge: Cambridge University Press.
Pylyshyn, Z. W. (1984) *Computation and Cognition.* Cambridge (MA): MIT Press.
Quine, W. V. (1969) *Ontological Relativity and Other Essays.* New York: Columbia University Press.
Rumelhart, D., and McClelland, J. L. (1986, ed.) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition.* Vol. I. Cambridge (MA): MIT Press.
Ryle, G. (1949) *The Concept of Mind.* London: Hutchinson University Press.
Schank, R. C. (1975) *Conceptual Information Processing.* New York: Elsevier.
Schank, R. C., and Abelson, R. (1977) *Scripts, Plans, Goals, and Understanding.* Hillsdale (NJ): Erlbaum Association.
Searle, J. (1980) Minds, brains, and programs. *The Behavioral and Brain Sciences* 3, 417–424.
Searle, J. (1992) *The Rediscovery of the Mind.* Cambridge (MA): MIT Press.

Skinner, B. F. (1953) *Science and Human Behavior.* New York: Norton.
Smith, J. C. (1990, ed.) *Historical Foundations of Cognitive Science.* Dordrecht: Kluwer.
Tomasello, M. (1999) *The Cultural Origins of Human Cognition.* Cambridge (MA): Harvard University Press.
Tomasello, M. (2009) *Why Do We Cooperate?* Cambridge (MA): MIT Press.
Tooby, J., and Cosmides, L. (1992) Psychological foundations of culture. In: J. H. Barkow, L. Cosmides, and J. Tooby (eds.) *The Adapted Mind.* New York: Oxford University Press.
Ullman, M. T. (2001). A neurocognitive perspective on language: The declarative/procedural model. *Nature Reviews Neuroscience* 2, 717–726.
Ullman, M. T. (2004) Contributions of memory circuits to language: The declarative/procedural model. *Cognition* 92, 231–270.
Vygotsky, L. S. (1978) *Mind in Society.* Cambridge (MA): Harvard University Press.
Wiener, N. (1948) Cybernetics or Control and Communication in the Animal and the Machine. Cambridge (MA): MIT Press.
Wiener, N. (1950) The Human Use of Human Beings: Cybernetics and Society. Boston: Houghton Mifflin.
Wilson, R. A., and Keil, F. C. (1999, eds.) *The MIT Encyclopedia of the Cognitive Sciences.* Cambridge (MA): MIT Press.

TOWARDS A NEW PHILOSOPHICAL PERSPECTIVE ON THE HISTORY OF COGNITIVE SCIENCE

Lilia Gurova

There are different ways to approach the history of cognitive science from a philosophical perspective and each of these approaches is based on a different assumption about the role which philosophy has played both in the birth and in the consequent growth of cognitive science.

1) One may start, for example, from Howard Gardner's assumption that cognitive science was born with a philosophical agenda (Gardner 1985), and then try to track what happened with this agenda in the following years.
2) Another option is to take a Popperian stance to the history of cognitive science assuming that each scientific research program is based on metaphysical assumptions, and then try to
   a) explicate the philosophical cores of the main research programs in cognitive science; and
   b) trace the changes in thus explicated philosophical cores.
   This is the line, for example, which Jean-Pierre Dupuy (2009) has declared to follow.
3) Contrary to (1) and (2), one may stay skeptical about whether either the research agenda, or the hard cores of the research programs in cognitive science are literally philosophical. But still one could agree that the scientific research agendas and the underlying assumptions are part of, and thus in a sense determined by, a larger intellectual milieu of which philosophy is an indispensable part. To view the history of cognitive science from this philosophical perspective means to trace the mutual influences between philosophy and cognitive science taken as relatively autonomous intellectual endeavors. The two-volume *magnum opus* of Margaret Boden (2006) is probably the best example of this approach.
4) Finally, one can build his/her philosophical account of the history of cognitive science by starting from the fact that philosophy has been recognized as one of the cognitive disciplines from the very beginning of cognitive science. Such a philosophical approach to the history of cognitive science is expected to focus on the place which philosophy has taken among the other cognitive disciplines, and on its, possibly changing, role in the common enterprise. An implementation of this approach can be found in the brief overview provided in Bechtel (2010).

Although starting from different assumptions, the above-listed four stances to the philosophical history of cognitive science are not incompatible. In fact, parts of the existing historical accounts successfully combine some of these seemingly different approaches. Thus,

Gardner, who has taken as a primary stance that cognitive science in itself is "deeply rooted in philosophy" (Gardner 1985: xiii), at the same time discusses extensively the role of the external philosophical context in the years immediately preceding the institutional birth of cognitive science. In a similar manner, Margaret Boden who has not made the strong claim that cognitive science literally follows a philosophical agenda, and who has spoken about the role of philosophy in cognitive science in terms of philosophy's influence on cognitive science matters, still admits that a plenty of philosophical issues arise inside the field of cognitive science, which the working cognitive scientists cannot simply ignore (Boden 2006: 1337).

It is important to stress that the assumptions that underlie the above-outlined philosophical perspectives on the history of cognitive science function as organizing principles of historical facts. Using different organizing principles leads to representing certain facts as more salient (or historically more important), while other facts are left less visible or even ignored.

Thus, when one starts from assumption (4) that philosophy as one of the cognitive disciplines is expected to contribute directly to the implementation of the cognitive science research agenda, he/she is in the first place occupied with looking for evidence for such direct contributions. Following this line, it will be natural for him/her to focus above all on the philosophical papers published in the main journals and conference proceedings, to take into account the percentage of philosophical papers and their topics, and to look for any tendencies in the change of both the percentage of published philosophical papers in the main journals and their topics. As we shall see, however, this strategy might lead (and in fact has led) to a certain perspective on the place of philosophy in the cognitive science enterprise, which is quite different from the view which one can obtain if he/she looks instead into the non-philosophical papers, paying attention, for example, to the kind of philosophical work the researchers without any background in philosophy have referred to.

In what follows, a small piece of research will be presented on the presence of philosophy in non-philosophical papers that have been published in the journal *Cognitive Science* in the last thirty years. As we shall see, the results of this research reveal a picture which is quite different from what seems to be the broadly shared view about how philosophy has influenced research in cognitive science since the 1980s, and about the kind of philosophy and philosophical topics that have been most influential in the field in this period. One cannot, of course, draw any definite conclusions on the basis of such a small portion of data. The data below should be rather read off as a support for the claim that a new philosophical perspective on the history of cognitive science is needed – a perspective which will take into account not only proper philosophical work done in and around cognitive science, but also the philosophical infusions which have penetrated the non-philosophical research of cognition.

Does the role of philosophy in cognitive science tend to decrease?

One can arrive at a positive answer to this question very easily if he/she takes into account the relative part of the philosophical papers published in the main journal of the Cognitive Science Society. As Figure 1 shows, the relative part of the papers published in *Cognitive Science* which have been written by philosophers has been decreasing since 1978. Can one read off this fact as evidence that the role of philosophy as such in cognitive science has been decreasing, too? This would be a quite hasty conclusion, as the following data

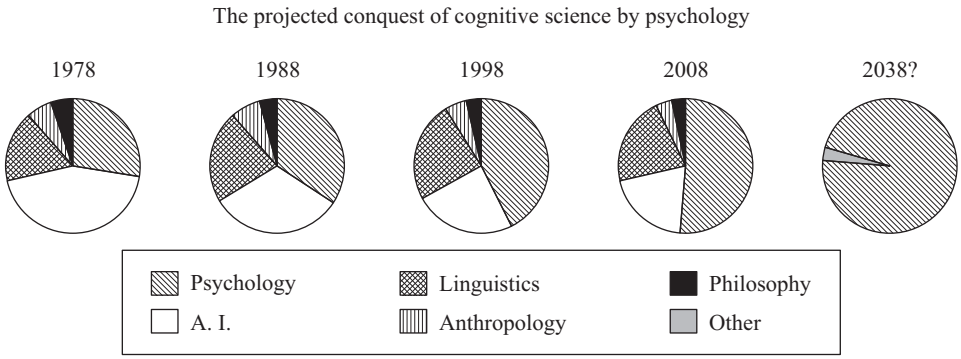


Figure 1. The proportion of authors by discipline in the first two issues of *Cognitive Science* in each decade, beginning in 1978 (taken from Gentner 2010: 330)

reveal. On Figure 2, one can see how the number of papers whose authors have referred to philosophy in one or another way has changed during the years. As we can see, the papers mentioning philosophy in some context almost doubled in the last decade. One may still stay skeptical about this tendency saying that the increase of the number of papers which simply

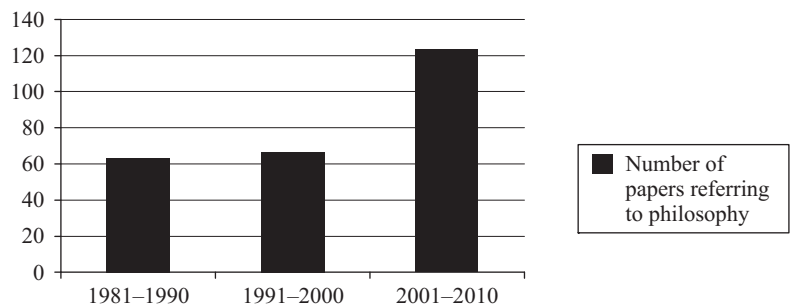


Figure 2. The number of papers published in *Cognitive Science* in 1981–1990, 1991–2000, and 2001–2010, which mention "philosophy" in their main text

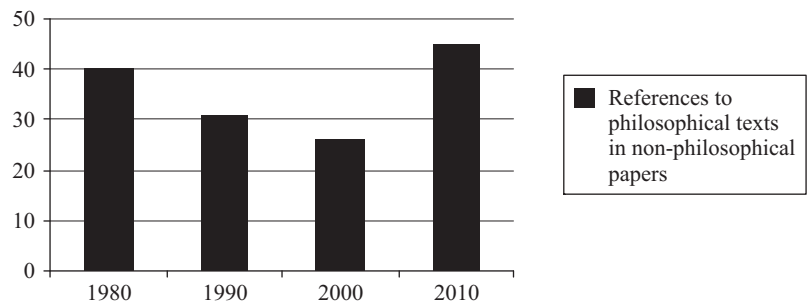


Figure 3. The number of references to philosophical texts in non-philosophical papers published in the first four issues of *Cognitive Science* for 1980, 1990, 2000, and 2010

mention philosophy does not tell us much about the increasing role of philosophy for non-philosophical research. However, Figure 3 shows the number of references to philosophical publications in the non-philosophical papers published in the first four issues of *Cognitive Science* for the years 1980, 1990, 2000, and 2010. As can be seen, there is no decrease in the number of references to philosophical publications. Taken together, the tendencies expressed on Figures 1, 2, and 3 rather tell us that, indeed, the role of philosophy in cognitive science has not been properly understood yet. Many philosophers of cognitive science today have arrived at the same conclusion – although for different reasons (see, for example, Brook 2009, Thagard 2009, Bechtel 2010).

Is the philosophy of mind the philosophical discipline which is most useful (or informative) for cognitive scientists?

When speaking about the philosophical disciplines which bear most on cognitive science, philosophers usually start with the philosophy of mind. In her discussion of the philosophical context, which gave birth to the core ideas of cognitive science, M. Boden, for example, wrote the following: "A fortiori, cognitive science is related to the philosophy of mind" (Boden 2006: 1335). In a chapter devoted to "the most basic philosophical issues that arise in and around cognitive science" (Harman 1993: 831), written by G. Harman for Michael Posner's collection of essays on the foundations of cognitive science (see Posner 1993), the three "most basic philosophical issues" discussed by Harman are issues central to the philosophy of mind: the mind–body problem, intentionality, and *qualia.* Two of the three philosophical issues "that bear on cognitive science" according to W. Bechtel (see Bechtel 2010) are also issues that are central for the philosophy of mind (the mind–body problem, and the nature of mental representations). Indeed, Bechtel's third issue that he insists is important for cognitive science (explanations in cognitive science) belongs to the philosophy of science. However, most philosophers of cognitive science who have accepted Andrew Brook's convenient distinction between the "philosophy in" and "philosophy of" cognitive science (Brook 2009) would locate this topic rather within the philosophy "of" cognitive science than within the philosophy "in" cognitive science. According to Brook, "philosophy in" is the part of

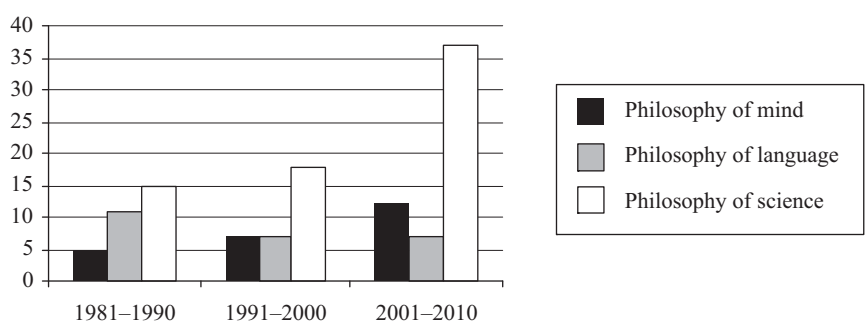


Figure 4. The number of papers referring to "philosophy of mind", "philosophy of language", and "philosophy of science" published in *Cognitive Science* in the periods 1981–1990, 1991–2000, and 2001–2010

philosophy which contributes to the immediate study of cognition. He insists (and many seem to agree with him) that the philosophies of mind and language are those that usually appear as "philosophy in" cognitive research. Philosophy of science bears on cognitive science, Brook suggests, only as "philosophy of" cognitive science, i.e., as an external critical reflection on the main methods and concepts which cognitive scientists have used in their research. Given these clarifications, one may expect, as Dennett (2009) does for example, that cognitive scientists would appreciate more "philosophy in", or the philosophy that immediately address the core issues of mind and language. The word search in the archive of *Cognitive Science* for the period 1977–2010 reveals, however, that as a philosophical discipline, "philosophy of science" is mentioned twice more often than "philosophy of mind", and that the interest in it seems to increase faster than the interest in the philosophy of mind, as Figure 4 shows. How should one understand this fact? Before suggesting an answer to this question let's see what the attitude of the non-philosophers in cognitive science is toward the central problem in the philosophy of mind: the mind–body problem.

Is the mind–body problem the most important philosophical issue arising in and around cognitive science?

Most philosophers, writing on philosophical issues of cognitive science, do seem to think so. The already mentioned paper of Gilbert Harman (Harman 1993) begins with a discussion of the main solutions to the mind–body problem that seem relevant to contemporary research in cognitive science. In the same vein, Margaret Boden claims that cognitive science owes its attraction largely to "its promise to help solve one of the greatest philosophical puzzles of all: the mind–body problem" (Boden 2006: 1337). In a recent paper W. Bechtel focused on three philosophical topics "that bear on cognitive science" (Bechtel 2010: 359), and the first of these topics is again the notorious mind–body problem.

Do non-philosophers in cognitive science appreciate the mind–body problem in the same manner? There is no evidence for that. The search done in all issues of *Cognitive Science* reveals that for the whole period between 1977 and 2010 the mind–body problem was mentioned explicitly only in 12 papers. Respectively, functionalism, which some philosophers of cognitive science have recognized as "the official philosophy of mind of cognitive science (Brook 2009: 217), was mentioned explicitly in fifteen papers only (see Figure 5). At the same time, "causation", which, to the best of my knowledge, has never been distinguished as one of the central philosophical issues arising "in and around cognitive science", has been discussed by non-philosophers much more often (see Figure 5 again). It is true that philosophers like Bechtel have demonstrated awareness of the fact that the cognitive scientists who embrace positions which philosophers of mind recognize as "functionalism", or "mind–brain identity", do not care about the philosophical discussions for and against these positions (Bechtel 2010). Bechtel, however, still believes that the methodological discussions among non-philosophers in cognitive science on how cognitive mechanisms ground in neural mechanisms could benefit from the parallel philosophical discussions on the mind–body problem. The reality, however, does not support his belief. The non-philosophers in cognitive science who have to take a position in relation to how the mind resides in brain structures simply assume such a position without showing the slightest willingness to get involved in any discus-

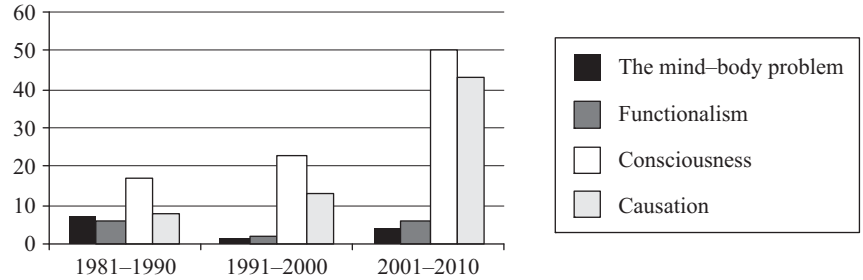


Figure 5. The number of references to "the mind–body problem", "functionalism", "consciousness", and "causation" in all issues of *Cognitive Science* published in the periods 1981–1990, 1991–2000, and 2001–2010

sion about whether their position is philosophically tenable. The assumption, for example, that cognitive processes "are localizable to some pattern of brain activity", is often treated as "given" (see Lenartowicz et al. 2010: 679).

Causation as a neglected topic by philosophers "in and around" cognitive science

Causation, causal inference, and causal explanations have always been central topics in philosophy of science. As it was already stressed, however, the shared view among the current philosophers of cognitive science is that, although important, the philosophy of science can only play a role as philosophy "of" but not as philosophy "in" cognitive research. According to the same shared view, central for philosophy "in" cognitive science are exclusively mind- and language-related topics such as concepts, meaning and reference, color vision, and consciousness (see Brook 2009). A central issue for the philosophy "of" cognitive science is the problem of explanation (Brook 2009; Bechtel 2010). The problems of causation and causal inference have not been considered as important neither for the philosophy "in", nor for the philosophy "of" cognitive science. Ironically, however, causation seems to be one of the topics which non-philosophers in cognitive science readily admit as needing philosophical clarification. Three of the eleven papers published in the first four issues of Cognitive Science for 2010 which refer to philosophical texts are related in some way or other to the problem of causation. All three of them (Lucas et al. 2010, Kushnir et al. 2010, Rips 2010) have cited extensively philosophical texts ranging from the classical Hume's treatise *An Enquiry Concerning Human Understanding* (1748) to some contemporary philosophical classics such as Frank Jackson's *Causal Theory of Counterfactuals* (1977), David Lewis' *Causation as Influence* (2000), Spirtes, Glymour and Scheines' *Causation, Prediction, and Search* (1993), and Woodward's *Interventionist Theories of Causation* (2007).

The latter names as well as their papers and books, however, are as a rule omitted from the recent accounts of the philosophy "in and around" cognitive science – which is another good reason to appeal for a more empirically informed research on what non-philosophers in cognitive science find interesting and helpful in both the historical and contemporary philosophical context. There might be other surprises for those who have accepted the shared view about

what philosophy "in" cognitive science is about, as the diagrams, representing the number of citations of non-living philosophers in all issues of *Cognitive Science* show (see Figures 6, 7, and 8). Among other things, the fact that Kuhn, Quine, and Wittgenstein are twice more often cited than any other non-living philosopher is in need of a thorough explanation.

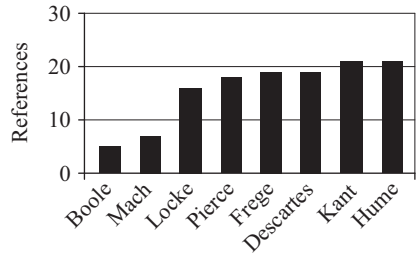


Figure 6. The number of references to philosophers who lived in the 17th–19th centuries in all issues of *Cognitive Science* (1977–2010)

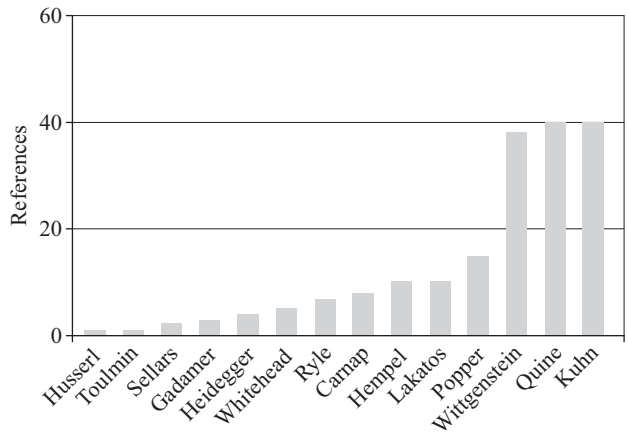


Figure 7. The number of references to non-living philosophers who worked in the 20th century in all issues of *Cognitive Science* (1977–2010)

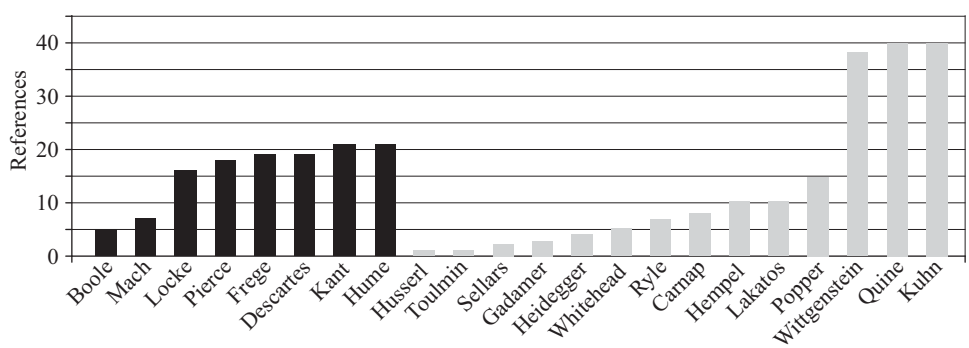


Figure 8. Here, the references from Figures 6 and 7 represented together: it is evident that the three non-living philosophers which have been most cited in *Cognitive Science* are Kuhn, Quine, and Wittgenstein

References

Bechtel, W. (2010) How can philosophy be a true cognitive science discipline? *Topics in Cognitive Science* 2 (3), 357–366.

Boden, M. (2006) *Mind as Machine: A History of Cognitive Science.* Oxford: Oxford University Press.

Brook, A. (2009) Topic introduction: Philosophy in and philosophy of cognitive science. *Topics in Cognitive Science* 1 (2), 216–230.

Dennett, D. (2009) The part of cognitive science that is philosophy. *Topics in Cognitive Science* 1 (2), 231–236.

Dupuy, J.-P. (2009) *On the Origins of Cognitive Science.* Cambridge (MA): MIT Press.

Gardner, H. (1985) The Mind's New Science: A History of the Cognitive Revolution. New York: Basic Books.

Gentner, D. (2010) Psychology in cognitive science: 1978–2038. *Topics in Cognitive Science* 2 (3), 328–344.

Harman, G. (1993) Some philosophical issues in cognitive science: Qualia, intentionality, and the mind–body problem. In: Posner, M. (ed.) *Foundations of Cognitive Science.* Cambridge (MA): MIT Press, 831–848.

Hume, D. (1748) *An Enquiry Concerning Human Understanding.* Indianapolis (IN): Hackett.

Jackson, F. (1977) A causal theory of counterfactuals. *Australasian Journal of Philosophy,* 55, 3–21.

Kushnir, T., Gopnik, A., Lucas, C., and Schulz, L. (2010) Inferring hidden causal structure. *Cognitive Science* 34 (1), 148–160.

Lenartowicz, A., Kalar, D., Congdon, E., and Poldrack, R. (2010) Towards an ontology of cognitive control. *Topics in Cognitive Science* 2 (4), 678–692.

Lewis, D. (2000) Causation as influence. *Journal of Philosophy,* 97, 182–197.

Lucas, C., Griffiths, T. (2010) Learning the form of causal relationships using hierarchical Bayesian models. *Cognitive Science* 34 (1), 113–147.

Posner, M. (ed.) (1993) *Foundations of Cognitive Science.* Cambridge (MA): MIT Press.

Rips, L. (2010) Two causal theories of counterfactual conditionals. *Cognitive Science* 34 (2), 175–221.

Spirtes, P., Glymour, C., and Scheines, R. (1993) *Causation, Prediction, and Search.* New York: Springer Verlag.

Thagard, P. (2009) Why cognitive science needs philosophy and vice versa. *Topics in Cognitive Science* 1 (2), 237–254.

Woodward, J. (2007) Interventionist theories of causation in psychological perspective. In: A. Gopnik, and L. Schulz (eds.) *Causal Learning.* Oxford: Oxford University Press, 19–36.

THE PREHISTORY: THE BIRTH OF COGNITIVE SCIENCE

PREHISTORY OF COGNITIVE SCIENCE: AN INTRODUCTION[1]

Andrew Brook

For the purposes of this chapter, the prehistory of cognitive science is the period up to about 1900. There then followed an interregnum during which cognition was not much studied, and the period of cognitive science proper followed that, starting in the 1950s and 1960s. The period up to about 1900 is called 'prehistory' because while philosophers and psychologists certainly studied the mind in this period, few envisaged a *science* of the mind. David Hume was one of the few exceptions and even he was only a partial exception because, while he hoped to build a scientific *model* of the mind, he did not do any scientific *investigation* of the mind. The prehistory came to an end when a group, a remarkably diverse group, of theorists began to do just that: Wundt, James, and Freud in particular. With the exception of Aristotle, all the figures that I will introduce in this short chapter came in or after the period, often called the early modern period, which is usually dated from about 1600.

The interregnum was the period in which behaviourism reigned supreme in psychology and logical empiricism in philosophy. Then begins the period of the cognitive revolution (the *second* cognitive revolution, some say, as Descartes and colleagues launched the first one). The first glimmerings of the revolution can be variously dated: the development of programmable computers in the Bletchley code-breaking establishment in England, of which Alan Turing was a member during World War II (Turing wrote the seminal papers on computational theory that underpin the computational model of the mind before, during, and shortly after this period),[2] the famous Hixon Fund Conference, *Cerebral Mechanisms in Behavior,* at Caltech in 1948, the publication of Chomsky's *Aspects of Syntax* in 1956 (a work that revolutionized the way not only language but the whole of human cognition was conceived), and perhaps other dates. The year 1956 was also the year in which the idea of a unified,

[1] This chapter is derived from the introductory essay in Brook 2007. The volume contains new essays on each of the ten authors that I am discussing here. The contributors are: Stellan Ohlsson, Noam Chomsky, Marcelo Dascal, Anne Jaap Jacobson, Andrew Brook (who also wrote the Introduction), Don Ross, Arthur Blumenthal, Peter Simon, Tracy Henley, and Patricia Kitcher. The authors were asked to identify what was of permanent value in the figure on whom they wrote. The contributors are all active cognitive scientists. The result was a volume closer in topics and style to contemporary cognitive science than one usually finds in histories.

[2] For many years, Americans in particular claimed ENIAC (*E*lectronic *N*umerical *I*ntegrator *A*nd *C*omputer), built at the University of Pennsylvania in 1946, as the first programmable computer. We now know that it was not. The group at Bletchley had created Colossus and other computers a few years earlier in the course of developing tools to break Enigma and Ultimate, Germany's ultra-high-secret military codes. Some versions of Colossus were programmable. Being under a thirty- to fifty-year gag order, the Bletchley group could do nothing about this misinformation until recently. Indeed, few people knew much, if anything, about Colossus until recently.

multidisciplinary research programme into human and artificial cognition was first officially articulated, at the now-famous *Symposium on Information Processing* at MIT, September 10–12, 1956. Whatever, by the mid-1970s, the cognitive revolution had prevailed, and cognitive science was well-established. The Cognitive Science Society was formed in 1977, and *Cognitive Science,* the Journal of the Cognitive Science Society, was started in 1979.

Like many other scientists, cognitive scientists tend to know relatively little about the history of their own subject. Such lack of knowledge would be a pity anywhere but it is perhaps particularly unfortunate in a discipline such as cognitive science, where the history is so long and where so much in the conception of the mind of at least classical cognitive science was handed down to us from long-past predecessors and still governs our thinking without much critical assessment.

This chapter does not aim to be a complete account of work in the prehistorical period. Indeed, I will fall far short of discussing all the important contributors in this period. I will introduce what I take to be the ten most important figures but that involves leaving many other significant contributors out. Some of the significant people not discussed include Antoine Arnauld, logician extraordinaire and Descartes' great interlocutor, John Locke, the source of what came to be called 'British empiricism', Thomas Reid, Scottish contemporary of Kant's and advocate of common sense as a source of genuine knowledge, John Stuart Mill and Jeremy Bentham, worthy empiricist successors to Hume, Hermann von Helmholtz, successor to both Hume and Kant and teacher of Freud's teachers, and John Babbage, inventor of the first mechanical calculating machine. What I am interested in is figures who have had an enduring influence, who continued to be important well beyond their immediate period. To work!

The context

In the history of cognitive research, 1879 is often taken as a watershed year. Wundt is credited with creating the first-ever psychology laboratory in the Department of Philosophy of the University of Leipzig in that year. In fact, he had been doing psychological experiments for almost twenty years by 1879 but 1879 has stuck as the year in which experimental psychology began. Even though the research that Wundt did was not much like behavioural experimental work today, focussing on introspection as it did, he is credited with launching the experimental side of cognitive science, work that has dominated the discipline ever since.

Likewise, in 1879 Frege published *Begriffsschrift* [Conceptual Notation], the seminal work that presaged his epoch-making *Foundations of Arithmetic* (1884). Though badly received at the time, this work launched the formal apparatus of logic and semantics in the contemporary era, apparatus that continues to constitute the formal foundations of cognitive science, and also of logic and analytical philosophy of language, to this day. Prior to Frege, logic had hardly made any progress since the time of Aristotle and was most often done as a sideline when it was done at all. Kant is a good example: He simply took Aristotelian logic over lock, stock, and barrel, and proceeded from there. To be sure, there were exceptions. Medieval writers such as St. Anselm, William of Occam and Nicholas à Cusa did important work, though broadly within the Aristotelian tradition. Descartes, Arnauld, and the Port Royal logicians also did work of enduring value at the intersection of logic and language, though it was lost again soon after. Leibniz formulated a grand research programme for logic, one, indeed, that Frege sometimes

saw himself as carrying out. One might also mention Boole (of Boolean algebra fame), though he worked only a couple of decades before Frege. Nevertheless, in general it is true to say that there were few major developments in logic from Aristotle to Frege.

The last decades of the 19th century were a period of considerable ferment in cognitive research. In these years, psychology, linguistics, logic, and semantic theory developed into distinct intellectual enterprises. Prior to then, such work as had been done in these areas had been done mainly by researchers who identified themselves as philosophers, or philosopher psychologists. Even the separation of psychology from philosophy occurred only in the 19th century – Wundt for example was a member of a Department of Philosophy, not Psychology. Prior to this time, cognitive theorists of all stripes called themselves philosophers. (Of course, so did a lot of other researchers. Sir Isaac Newton's Chair at Cambridge, for example, is called the Chair in Natural Philosophy to this day, and the term 'PhD' is short for 'Doctor of *Philosophy*' in Latin.) By shortly after 1900, psychology had decisively separated from philosophy, linguistics had come into its own as a separate discipline (though work on language did not have much influence on general cognitive research until Chomsky and the 1950s – Whorf's work in Harvard's famous five-cultures study of the 1950s[3] was one of rare exceptions), and the logical/semantic tradition that Frege's work made possible had established itself via Russell, the early Wittgenstein, and the logical positivists of the Vienna School as a dominant influence within English-speaking philosophy.

Ten major contributors in the prehistory of cognitive science

In this chapter, I will, as I said, introduce ten major figures from the prehistory. They are: Aristotle, Descartes, Hobbes, Hume, Kant, Darwin, Wundt, Frege, James, and Freud. These ten researchers discuss an enormously diverse range of topics in an enormously diverse range of ways. In my view, this diversity is significant; indeed, it is still a feature of cognitive science. If cognitive science is unified as a conception, it is much less unified as an activity. A great many voices and a great many topics contend with one another, voices ranging from hard empirical and computational modelling at one end to broad speculations about situated cognition and chaotic systems at the other, topics ranging from 'classical' ones such as syntax, lexical processing, perception, and reasoning systems to such things as connectionism, dynamic systems, and cognitive neuroscience.4 We find much the same diversity in the major figures of the prehistory.

Aristotle

The story of serious, systematic thinking about cognition goes back as far as Aristotle, indeed perhaps even to Plato (see Table 1 for dates). Aristotle, for example, articulated a distinction

[3] In what is widely considered to be the first major empirical project in anthropology, a group of Harvard anthropologists and linguists studied five linguistic/cultural groups in the four-corner area where Arizona, Utah, New Mexico, and Colorado meet. The groups were Hopi, Apache, Navaho, English-speaking ranchers, and Hispanics.

[4] Sometimes one finds both sides in one person. William Clancey is a good example. Compare his book *Situated Cognition* (1996) with his earlier AI work developing expert systems such as MYCIN.

between practical and theoretical reason that is still accepted and continues to be influential. Theoretical reasoning is reasoning about what to believe, what is the case, etc., while practical reasoning is reasoning about what to do, what ought to be the case, etc. Aristotle saw so far into the distinction that he even connected it to the distinction between teleological and mechanistic explanations. Practical reasoning concerns what he called final causes, that is to say, goals and purposes. Theoretical reasoning, though concerned with final causes, also concerns itself with what he called efficient causes or what we now would call mechanisms and causes, period. Aristotle articulated a system of sentential logic that survived unscathed until the time of Frege. And his account of what we would now call human cognition in *De Anima* is the first attempt ever to give something like a systematic description of human cognition.

Table 1. Fourteen main figures in the prehistory of cognitive science

Plato (438 BC–347 BC)	
Aristotle (384 BC–322 BC)	
Descartes (1596–1650)	
Hobbes (1588–1679)	
Empiricism	**Rationalism**
Locke (1632–1704)	Spinoza (1632–1677)
Hume (1711–1776)	Leibniz (1646–1716)
Kant (1724–1804)	
Darwin (1809–1882)	
Wundt (1832–1920)	
Frege (1848–1925)	
James (1842–1910)	
Freud (1856–1939)	

Nevertheless, in important respects the story of *cognitive* research begins later. Aristotle described something recognizable as cognitive functions, indeed saw them as functions of the body and arguably as biological functions, but he had no conception of representation as we now understand it, nor of consciousness, nor of memory, nor of perception as the processing of information in the brain, nor..., nor..., nor... That he did not have a concept of a mental representation, a concept, that is to say, of something that functions by standing for or referring to something else, is arguably the thing that most centrally separates his work from all work on cognition in the modern era, which begins about 1600.[5] Aristotle held that perception, for

[5] There is an interesting and complicated story to be told about the history of the notion of a representation. While there is not much evidence of any clear notion of a *mental* representation prior to the time of Descartes and Hobbes, something like our current notion of a *linguistic* representation, a sign, goes all the way back to the Stoics (Dascal and Dutz 1997). Why one notion of something presenting or standing for something else should have developed so much sooner than the other is an interesting question.

example, consists roughly in taking the essential structure of the perceived object into the mind. Thus perception is 'built out of' aspects of the thing perceived, not out of states and processes that *represent or stand for* the thing perceived. Aristotle's view of perception was the view adopted (or assumed) by most theorists, both Platonists and Aristotelians, until the modern era. St. Thomas Aquinas is a good example.

Though there are anticipations of our current conception of a representation in some late medieval thinkers (Pasnau 1997), it achieved its current form at about the time of Descartes and Hobbes (see Table 1 for dates). It is with this conception that the study of cognition as we now understand it really begins (though the idea of a *science* of cognition was still centuries away even at that point). It is not that nothing happened in the roughly two thousand years between Aristotle and Descartes. In fact, a great deal happened, more than is usually realized. In late Roman times, for example, St. Augustine had already articulated the inference for which Descartes is famous, the inference *cogito ergo sum* ('I think therefore I am'). But Augustine had no precise conception of *what* the 'I' is, indeed he probably did not get much further than Aristotle on that score. (NB. I am painting with a very broad brush here.) Descartes did, and so did Hobbes.

Descartes

Descartes held that the mind is 'a thing that thinks'. What he meant by 'thinks' was something very different from what Aristotle would have had in mind. Descartes conceived of the materials of thinking as representations in the contemporary sense. And Hobbes was the first to clearly articulate the idea that thinking is operations performed on representations. Here we have two of the dominating ideas underlying all subsequent cognitive thought: the mind contains – and is a system for manipulating – representations.

Descartes' contribution to our conception of human cognition was massive. The central aspects of it endured with no serious competitors until about the late 1950s. These aspects include:

1) The notion of a representation, i.e., something cognitive that stands for something else.
2) The idea that representations are in the head (in the mind).
3) The idea of the mind as a unified system of representations, a unified being to whom representations represent.

Dennett (1991) calls the last notion, the idea of the mind as a being to whom representations represent in a kind of quasi-spatial arena inside the head, the Cartesian Theatre. All these ideas endure to this day. They all figure, for example, in Fodor's representational theory of mind. Many of the most persistent problems about cognition also stem from them, e.g., the problem of knowledge of the external world and of other minds. Many recent developments in cognitive thinking are direct reactions to them, e.g., Gibson's ecological cognition (Gibson 1979) and externalism about the content of representations (the claim that 'meanings just ain't in the head' in Putnam's memorable 1975 phrase). In addition to Gibson, other serious alternatives to the Cartesian picture as a whole since World War II have included behaviourism (Skinner 1974, Ryle 1949), Dennett's multiple-drafts alternative to the Cartesian Theatre

(1991, for commentary see Brook and Ross 2002), and connectionism and neurophilosophy (P. M. Churchland 1984, 1994; P. S. Churchland 1986). However, the Cartesian picture remains overwhelmingly the dominant picture in cognitive science.

Descartes' model of the mind has been extensively discussed in recent decades, by Ryle (1949) with his critique of the 'ghost in the machine', and Dennett (1991) with his critique of the Cartesian theatre, so I won't say anything further about it. In addition to articulating a representational model of the mind, Descartes is generally credited with being the father of rationalism, the view that knowledge or at least some forms of knowledge can be achieved independently of experience. His view remains alive to this date, for example in the Chomskyan claim that much of our knowledge of grammar is innate.

Against the empiricist-sounding dictum of Aristotle that "nothing is in the mind that is not first in the senses", Descartes argued that what the mind achieves by *reflection* on things is closer to knowledge of their nature than what it *observes* about them.[6] Descartes was thus the first to pay serious attention to the balance between the role played by the mind and the role played by sensible experience in the acquisition of knowledge. In the form of the battle between empiricism and rationalism, this problem achieved its first resolution only with Kant and continues to be a live issue today, for example in the controversies over the size of the mind's 'top-down' contribution to abstracting patterns from sensible stimulations.

Nor was Descartes' originality limited to the mind. With Galileo and others, he was also one of the originators of the mechanistic conception of the universe, for example, and he did extensive experimental neurophysiology. Indeed, he laid down a neurophysiological conception of vision and of cognition more generally.[7] Descartes made language a central indicator of the presence of a mind (though, as Dascal shows in Brook [2007], he also *separated* language from cognitive activity more radically than most would now). He had a major influence on the work on logic and language of his contemporaries at Port Royal. (Many important ideas of the latter group were lost again.) All of the ideas introduced in the preceding paragraphs are of enduring value.

To be sure, not everything that Descartes believed about the mind has lived on. For modern tastes, he placed the balance between the contribution of the mind and the world too far on the mind side, being the good rationalist that he was. Rationalism as exemplified by theorists such as Spinoza and Leibniz is the view that the representations to be trusted are the ones arrived at entirely inside the head by processes of reasoning alone. (The crucial kind of reasoning here is the exploring of the semantic implications of one's concepts and propositions.) When cast less austerely and therefore more plausibly, a modest form of rationalism lives on, as we said, in Chomsky's conception of universal, innate grammar, and also in Fodor's language of thought hypothesis (the latter is the view that the materials out of which our concepts

[6] Aristotle's dictum might appear to be ultra-empiricist but we should be cautious about jumping to this conclusion. His picture of perception was utterly different from ours: he thought that essential features of the structure of objects literally move from the objects into us. His picture being so different, nothing but confusion is likely to result from giving the two conceptions of the source of knowledge the same name.

[7] In connection with this, Descartes and his tradition achieved the first clear articulation of the problem of unifying knowledge formulated at different levels and in different vocabularies, goal-directed language vs. mechanistic language, for example. Aristotle had anticipated some aspects of the issue but Descartes confronted it head-on. It is a major issue today.

are constructed are also universal and innate). However, few theorists now would push the idea as far as Descartes seems to have done.

More importantly, strongly impressed by the complexity of language and the free creativity that it made possible, Descartes held that minds able to use language are things entirely 'separate and apart' from the body, non-spatial, non-material entities made up of who knows what, and, together with this, that non-human animals do not have minds. (He held, for example, that non-human animals cannot feel pain and dissected them without anaesthetic.) Few contemporary cognitive scientists follow him in any of this. That said, his explanatory dualism persists. Many cognitive scientists think that we are permanently stuck with a duality of explanations – explanations of neurological processes in the language of the neurosciences, and explanations of cognitive function in the language of folk psychology or some other teleological language. (This is one way in which Aristotle's distinction between theoretical and practical reasoning lives on.) Some cognitive scientists even think that we must retain a dualism of properties, e.g., between neurological composition and cognitive functioning, or between cognitive functioning and qualitative feel, 'qualia' in philosophers' jargon (Chalmers 1996). But few now accept Descartes' ontological dualism, however obvious it seemed to him; few now think of a person as a 'union' of two utterly different kinds of thing. Indeed, the reverse seems obvious to most people.

Hobbes

In this, Hobbes has had the more enduring influence. Hobbes was a near-contemporary of Descartes', indeed wrote the best-known of the six series of objections to Descartes' *Meditations* of 1645. On the fundamental nature of the mind, Hobbes and Descartes utterly disagreed. Hobbes urged that the mind simply is the brain, or certain aspects of it. This is connected to his single greatest contribution to our conception of cognition, the idea, as he put it, that "all reasoning is but reckoning" (1651 I/5: 1–2) – all thinking is computation. Put Hobbes' mechanistic materialism together with Descartes' notion of representation, and you have the fundamentals of the contemporary picture of cognition: cognition consists of computations over representations.

If Hobbes was a materialist, there is a good deal more to his view of knowledge than simple empiricism. Hobbes' claims about the tight relationship between language and thought are closer to the spirit of rationalism, to Spinoza and Leibniz, than to empiricists such as Locke and Hume. Indeed, Hobbes inspired a research programme on thinking that was at the centre of both rationalism and empiricism in the 17th and 18th centuries, a programme that continued at least as far as Stewart in the 19th century. On the other hand, though Descartes is supposed to be the father of rationalism, his separation of language and thought was much more in the spirit of empiricism than of rationalism. Indeed, both Hobbes and Descartes cross-cut the time-worn division of early modern cognitive thinking into empiricism and rationalism.

Hume

As we have indicated, rationalism of one kind or another was one of the great stances on knowledge acquisition and validation of the early modern period. The other was empiricism, as in the Aristotelian dictum that "nothing is in the mind that is not first in the senses" interpreted as we would now interpret it (see note 6). The British philosopher John Locke is generally viewed as the originating figure of what came to be called British empiricism. However, there is at least some ambiguity about the extent of Locke's empiricism. By contract, the Scottish philosopher David Hume was unambiguously and radically an empiricist. He carried out the empiricist programme more comprehensively and rigorously than anyone before him (and maybe since). Hume held that there is no source of knowledge except sense experience. He also held that an empiricism rigorously followed out will end up denying that sensible experience has anything like the structure of a language – sensible representations are like pictures, not propositional structures, and associations govern their relationships, not propositional relations. This is enough by itself to make Hume the grandfather of behaviourism and of connectionism. Hume also saw a set of sceptical problems as lying at the heart of empiricism. According to him, we can never justify our beliefs about the world external to us, the future, or even the past! However, how one views this sceptical streak in Hume's work depends very much on how one views his project as a whole. If one views him as holding to a picture of representations as like objects of some kind, then one must see him as mired in deep sceptical problems indeed. If one sees him as holding, in the spirit of later thinkers, that representations are cognitive acts of some kind, the issue about scepticism may take on quite a different cast.

Hume not only took empiricism about the contents of knowledge more seriously than anyone before him and maybe since, he also insisted that theories of mind stay within empiricist bounds. In particular, he insisted on what we would now call a naturalized epistemology. Not just the mechanisms by which we acquire knowledge but also *the standards by which we assess knowledge claims* have to be derived entirely from what nature provides. Likewise, by insisting even more rigidly than Descartes that everything about the content of representation is 'in the head', he formulated a picture of the content of representations that is still orthodoxy.

Now called *individualism,* it remains the view of most cognitive theorists even in the face of a recent challenge, *externalism.* Externalism is the view that the content of representations, what representations are about, consists of a relationship of some kind between what is going on in the head, and what is found in the world (Putnam 1975). Some (e.g., Clark and Chalmers 1998) even urge that some aspects of cognition lie outside the head. Externalism is largely confined to some philosophers of mind and has never had much influence in the rest of the cognitive community. (Some philosophers view J. J. Gibson and the more recent situated cognition movement as varieties of externalism but this view is disputable [Brook 2005].)

Kant

Kant brought empiricism and rationalism together. Gaps, as Kant saw it, in Hume's empiricism, and the sceptical problems about the nature of the self, and the knowledge that it seemed to Kant to entail aroused him from what he called his "dogmatic slumbers" – an uncritical submersion in the rationalism of his time. Spurred by the example of the trouble

that radical empiricism had caused Hume, Kant argued that the element in knowledge advocated by rationalism and the element advocated by empiricism are both necessary – to acquire knowledge, we need both sensible input and *a priori* activities of the mind. As he put it in a famous aphorism, "thoughts without content are empty, intuitions without concepts are blind" (1781: A51 = B75). That is to say, we cannot confirm or disconfirm conceptualizations without experiential evidence, no matter how carefully we think about the conceptualizations – but we cannot organize experience without applying concepts to it. The first cuts against full-blown rationalism, the second cuts against extreme forms of empiricism. Kant's resolution of the empiricist/rationalist tension is now widely accepted in cognitive science, some connectionists being among the few exceptions.

Kant's other views about the mind have also been incorporated into cognitive science: Kant's view about the mind as a system of functions, and his views about the right method to study the mind in particular. For these reasons, Kant can even be viewed as the grandfather of cognitive science.

If some of Kant's central ideas about the mind live on, it is interesting that a number of the ideas that he held most dear have played hardly any role in contemporary cognitive science at all. This is true of Kant's claims about the mind's synthesizing powers, about its various mental unities (in particular, the unity of consciousness), and about consciousness of self. Not only have these views not been superseded by cognitive science, they have never even been assimilated by it – and they deserve to be. Or so I have urged (Brook 1994).

19th century

If Wundt, Frege, and 1879 are the divide between the prehistory of cognitive science and the next period, Kant's *Critique of Pure Reason* of 1781 is the divide between the 18th and 19th centuries. Empiricism as Hume had laid it out in the *Treatise of Human Nature* (1739) continued to influence thought about cognition in the 19th century but the dominant influence, certainly in the German-speaking world, was Kant. Except for a few stubborn empiricists, work on cognition in German in the 19th century and even a lot of work in English consisted of spelling out and beginning to test ideas that Kant had articulated. Indeed, that is true of a lot of work on cognition up to the present. (The influential Cartesian ideas that we listed earlier live on in Kant's picture so Kant's picture continuing to have influence is also Descartes' picture continuing to have influence.) There have been movements that rejected the Kantian picture of cognition, of course. We have already mentioned connectionism and could add behaviourism (a form of extreme empiricism), though it was never part of cognitive science. But classical cognitive science and the great majority of cognitive researchers up to the present hold to a model of the mind that is Kantian in many essentials. Though, as we said, contemporary theorists have neglected some topics dear to Kant's heart, they conceive of the mind largely as Kant conceived of it.

The 19th century saw a blossoming of theorizing about cognition. A great deal of it did not add much that was really original to the two models that we were left with at the end of the 18th century – empiricism, and the Kantianism synthesis. J. S. Mill, Herbart, and Helmholtz might be considered examples. Among 19th-century figures who did add major new ideas, one thinks immediately of Darwin.

Darwin

Anticipations of Darwin's theory of evolution can be found earlier but as a well-articulated theory based on imposing and powerful evidence, Darwin's work had no antecedents. Evolutionary theory is coming to play a central role in cognitive science, and that for a variety of reasons.

First, evolutionary theory is an excellent way to approach the important task of reuniting cognitive theory and neuroscience. Cognitive theorizing and biology were deeply interanimating in Darwin's time but by the time of the great cognitive revolution of the 1970s, the two had come apart. Entranced by the computer metaphor, cognitive scientists of the classical period urged that, like computer system designers, we can understand the functioning, the 'software', of the mind/brain without needing to know much about how those functions are implemented in the brain. (Indeed, it was common to refer to the brain as the 'wetware'.)

Second, evolutionary theory is an excellent way to approach the task (and also, some think, the limitations) of building a purely naturalistic epistemology – an account of knowledge acquisition within the limits of what nature has provided (including acquisition of knowledge about the mind itself). It took a long time for evolutionary theory to come to play any important role in cognitive science but it is now the case that cognitive scientists who ignore Darwin do so at their peril.

Wundt

Alongside the idea that cognition has evolved, Wundt introduced a second element that was largely new: the idea that claims about cognition should be submitted to empirical test. Of course, the experimental method did not originate with Wundt. What he did was to find a way to apply it to claims about cognition. Few would disagree with Wundt about the importance of experimental verification today.

For Wundt, the experimental method was not an end in itself. Wundt thought that it revealed deep aspects of the mind that other methods do not reveal. Interestingly enough, for Wundt, the mind thus revealed fits Kant's picture better than the empiricist picture. However, Wundt's picture was not Kantian in every respect. His rejection of the idea of discrete, persisting representations resonates more with anti-representational views such as situated cognition than it does with Kant, for example.

Frege

A third new development in the 19th century from an entirely different direction was Frege's invention late in the century of the concepts and tools of modern symbolic logic and semantic theory. Indeed, most of the formal foundations of contemporary cognitive science were articulated by him. His work is the basis not just of logic and semantic theory but also of computational theory, which was the basis, in turn, of the computer revolution and artificial intelligence. Given that Frege himself fought ferociously to separate logic and semantic

theory from psychology, there is a certain irony in the fact that his work laid the foundations for the whole formal side of cognitive science.

Frege's many contributions defy brief summary (for details, see Peter Simon's excellent paper in Brook 2007). However, the formal analysis of language that he pioneered, his concept of a formal system with a rigorous syntax, his proof theory and semantics, and his approach via exact analysis to meaning, reference and thought in many of their forms are now indispensable to more formally-oriented cognitive science. Quite an accomplishment for a single thinker!

James

James' place in the prehistory of cognitive science is a bit different from Darwin's, Wundt's, or Frege's. With James there are no stunning new ideas that changed the shape of cognitive research forever. His body of work as a whole had a major influence but because of its accessibility and the breadth of issues that James took on, not because of major innovations. His best-known contributions are to a topic that slid from the view of most cognitive researchers not long after his time, and surfaced again, at least as an empirical enterprise (it had never disappeared in philosophy), as late as the 1980s – consciousness. Probably the single best-known concept in James is his notion of the stream of consciousness. James also made significant contributions to our understanding of association, memory, imagery, imagination, and reasoning.

Perhaps his most important contribution was his distinction between explanation by reference to biological foundations and explanation by reference to social/behavioural factors – a new version of the explanatory dualism that, as we saw earlier, goes all the way back to Aristotle. Finally, and reason enough by itself for including James as a central figure in the prehistory of cognitive science, his articulation of the idea of the mind as a system of functions is fuller than any in his time, indeed any prior to about thirty years ago.

James did not originate the functionalist conception of the mind. The basic idea goes back to Plato and Aristotle, and we find a full statement of what a mental function is and how the mind as system of functions works in Kant (Meerbote 1989; Brook 1994: Chapter 1). James carried an already-existing idea further. However, he carried it a great deal further. (There is a nice irony to James unwittingly following Kant in this way: James ridiculed Kant as few others have ever done, see Brook [1994: 1] for a leading example.)

Freud

Freud fits into the prehistory of cognitive science in yet another different way. He was a great innovator, of course: his theories of unconscious anxiety, defence transference, and so on, his view that neurotic and often psychotic emotion and cognition are meaningful, his idea that patients have reasons for feeling and thinking as they do, his theory that dreams contain crucial information about the dreamer, his claims that young children have a far richer, complicated, and structured sensuous life than anyone had realized, and the like changed the way we conceive of the mind. But Freud's ideas have not directly influenced cognitive science in

the way that Darwin's, Frege's, and Wundt's have. Indeed, far from being influenced by psychoanalytic theory, many cognitive scientists are deeply suspicious of it. His importance lies in a different direction.

What makes Freud important to cognitive science is that he was the first to build a comprehensive interdisciplinary model of the mind. Where he succeeded, and especially where he failed still have a great deal to teach us. Freud's fullest statement of his model (1895) was not published in his lifetime but the ideas in it shaped his thinking for the rest of his life. Perhaps the greatest problem facing such models is that they are hostage to the state of knowledge in the relevant disciplines at the time. Freud attempted to draw together in a single model everything significant known (or believed) about the mind by neurobiology, psychology, anthropology/archaeology, and evolutionary theory in his time. Unfortunately, this effort entails that if any of them were seriously wrong, his model was going to be seriously wrong, too. And the neurobiology of his day, built on a kind of hydraulic, reflex model of the forces of the mind/brain, was seriously wrong (to speak only of it). Interdisciplinary models of the mind in cognitive science continue to be hostage to surrounding science in exactly the same way.

Indeed, Freud may well have had more faith in the neurobiology of his time than did researchers in the field – something we still find today sometimes when cognitive scientists attempt to import chaos theory, or dynamic systems theory, or neural network architectures, or quantum mechanics into their work.

Conclusion

We will close where we began. No single discussion could come remotely close to doing justice to the range of ideas about cognition articulated prior to 1900. Nor was that my aim. Rather, I have tried to introduce ten researchers prior to 1900 whose ideas continue to shape how we conceive human cognition.

References

Brook, A. (1994) *Kant and the Mind.* New York, and Cambridge: Cambridge University Press.

Brook, A. (2005) My BlackBerry and me: Forever one or just friends? In: Nyíri, K. (ed.) *Mobile Understanding: The Epistemology of Ubiquitous Communication.* Vienna: Passagen Verlag, 55–66.

Brook, A. (ed.) (2007) *The Prehistory of Cognitive Science.* Basingstoke, UK: Palgrave Macmillan.

Brook, A., and Ross, D. (2002) *Daniel Dennett.* (Series: Contemporary Philosophy in Focus). New York: Cambridge University Press.

Chalmers, D. (1996) *The Conscious Mind.* Oxford: Oxford University Press.

Churchland, P. M. (1984) *Matter and Consciousness.* Cambridge (MA): MIT Press (1st edition: 1984; 2nd edition: 1988).

Churchland, P. M. (1994) *The Engine of Reason, the Seat of the Soul.* Cambridge (MA): MIT Press.

Churchland, P. S. (1986) *Neurophilosophy.* Cambridge (MA): MIT Press.

Clancey, W. (1996) *Situated Cognition.* New York: Cambridge University Press.

Clark, A., and D. Chalmers. (1998) The extended mind. *Analysis* 58, 7–19.

Dascal, M., and Dutz, K. (1997) Beginnings of scientific semiotics. In: R. Posner, K. Robering, and T. A. Sebeok (eds.) *Semiotics: A Handbook on the Sign-Theoretic Foundations of Nature and Culture.* Vol. 1. Berlin: W. De Gruyter, 746–762.

Dennett, D. C. (1991) *Consciousness Explained.* Boston: Little Brown.

Freud, S. (1895) *Project for a Scientific Psychology.* In: J. Strachey (trans. and ed.) *The Complete Psychological Works of Sigmund Freud.* Vol. 1. London: Hogarth Press, and the Institute of Psychoanalysis (1966).

Gibson, J. (1979) The Ecological Approach to Visual Perception. Boston: Houghton-Mifflin.

Hobbes, T. (1651) Leviathan or The Matter, Forme and Power of a Commonwealth Ecclesiastical and Civil. In: E. Curley (ed.) Indianapolis, IN: Hackett (1994).

Hume, D. (1739) *Treatise of Human Nature.* In: L. A. Selby-Bigge (ed.) Oxford University Press (1962).

Kant, I. (1781, 2nd ed.: 1787) *Critique of Pure Reason.* In: Norman Kemp Smith (trans.) *Immanuel Kant's Critique of Pure Reason.* London: Macmillan (1963).

Meerbote, R. (1989) Kant's functionalism. In: J. C. Smith (ed.) *Historical Foundations of Cogntiive Science.* Dordrecht: Reidel, 161–187.

Pasnau, R. (1997) *Theories of Cognition in the Later Middle Ages.* Cambridge, and New York: Cambridge University Press.

Putnam, H. (1975) The meaning of 'meaning'. *Mind, Language and Reality. Philosophical Papers.* Vol. 2. New York: Cambridge University Press, 215–271 (at p. 227).

Ryle, G. (1949) *The Concept of Mind.* New York: Barnes and Noble.

Skinner, B. F. (1974) *About Behaviorism.* New York: Random House.

EARLY APPARATUS-BASED EXPERIMENTAL PSYCHOLOGY, PRIMARILY AT WILHELM WUNDT'S LEIPZIG INSTITUTE[1]

H. Maximilian Wontorra

Despite a multitude of contributions to the historiography of early experimental psychology, this topic has rarely been dealt with from an apparatus-oriented perspective. Therefore, this chapter introduces the two most prominent research lines of early apparatus-based experimental psychology with the respective investigations, conducted primarily at Wilhelm Wundt's (1832–1920) Leipzig institute, the world's very first institute of experimental psychology. These two research lines are referred to as *mental chronometry* and *attempts to quantify the phenomena of consciousness*, respectively.

Mental chronometry

Mental chronometry was concerned with measuring the time consumed by basic mental operations such as stimulus discrimination, or choice of the adequate reaction to a certain stimulus. These chronometric investigations constituted the first coherent research program in experimental psychology. They were inspired, on the one hand, by inaccurate astronomical observations, and, on the other hand, by the determination of the unexpectedly slow signal propagation velocity in nerves by Hermann von Helmholtz (1821–1894) in the middle of the 19th century.[2] Helmholtz measured the respective velocities at about 30 meters per second in frog nerves and at 60 to 80 meters per second in human nerves (Helmholtz, 1850a, b; 1852).

As the astronomical problems mentioned above are closely related to what we nowadays call the paradigm of distributed attention, this topic will be shortly sketched. In early 1796, Sir Nevil Maskelyne (1732–1811), the Fifth Astronomer Royal at the Greenwich Observatory, dismissed his assistant David Kinnebrook (1772–1802) after only two years on duty. In the end, Kinnebrook's transit recordings of celestial objects diverged from Maskelyne's by about eight tenth of a second, for what Maskelyne had no other explanation than Kinne-

[1] This investigation was funded by the German Research Foundation (DFG, SCH 375/18–1).

[2] To determine the respective velocity in frog nerves, Helmholtz attached a frog muscle with the corresponding efferent nerve to a drawing device. He then stimulated the nerve electrically and produced plots of the muscle contraction over time. From the nerve's length and the latency of muscle contraction he calculated the propagation velocity. For human nerves, he determined this velocity by dermal electrical stimulation. This was a seminal work as only a couple of years earlier Johannes Peter Müller (1801–1857), another famous German physiologist of the 19th century, had expressed his conviction that this propagation velocity would forever exceed the range of measurability.

brook's carelessness. In those transit observations the observer simultaneously had to attend to the beats of a second signal and to the transit of the object through the meridian wire of the telescope. From the object's position at the last beat *before* transit and at the first beat *after* transit, the observer had to estimate transit time. Apparently, Maskelyne could not imagine that mental operations as listening to the beats of the clock and watching the celestial body on its trajectory are time-consuming competitive mental tasks.

In the early 19th century, the famous mathematician and astronomer Friedrich Wilhelm Bessel (1784–1846) made observations by means of a comparable method at the Königsberg Observatory in Prussia, and he found divergences of even more than 1 second between experienced observers (Bessel 1822). Only now was the scientific community sensitized to the time response of mental operations. While in the following years, astronomy eagerly tried to substitute the "inert system" of human information processing by technical devices, physiology and, modestly delayed, experimental psychology concentrated on investigating exactly these relatively inert mental processes quantitatively.

One important technical precondition for measuring these response times was an easy-to-use chronometer with a high resolution, down to the millisecond, if possible. In principle, this instrument was available in Charles Wheatstone's (1802–1872) so-called "chronoscope" from the 1840s. Initially conceived for technical time measurements, chronoscopes (see Figure 1) were clockworks driven by a heavy weight. These devices were started or stopped by closing or opening an electrical circuit into which they were integrated. Wheatstone's chrono-

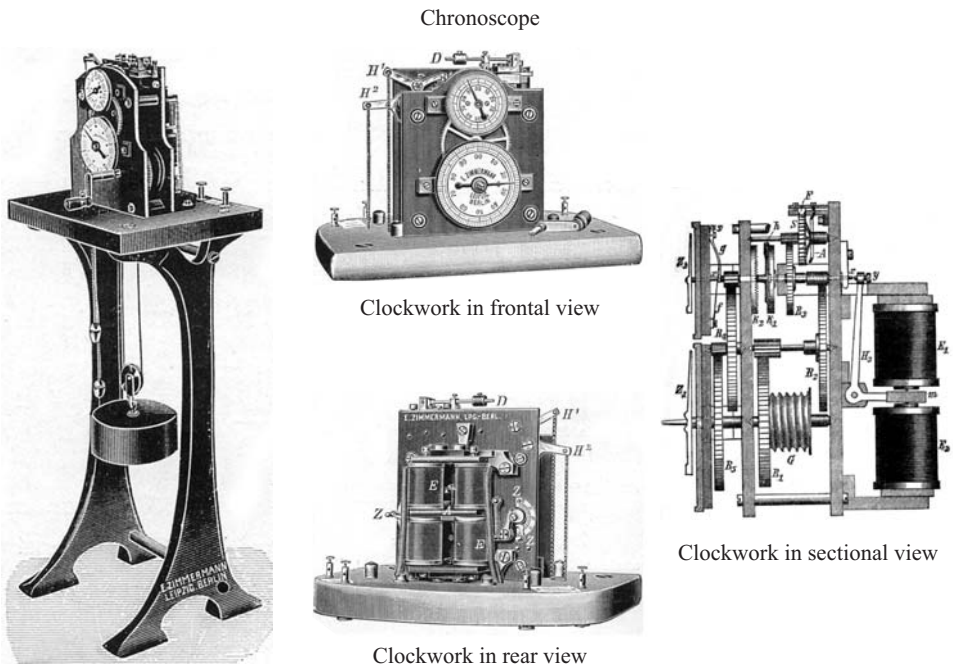


Figure 1. The Hipp-type chronoscope. For a detailed explanation, please see text. (Sources: Zimmermann 1928: 110, and Wundt 1908–11, 3rd vol.: 369, respectively)

scope was essentially enhanced by the German clockmaker Matthias Hipp (1813–1893) in the 1860s by separating the hands of the clock from the clockwork itself. As a result, the inertia of the system was immensely reduced. This brought about a drastic improvement to Wheatstone's chronoscope, in which the complete clockwork had to be started at the beginning, and it had to be stopped at the end of the time interval to be measured. In Hipp-type chronoscopes, the hands were connected to the permanently running clockwork at the beginning of the respective time interval, and they were disconnected at the end of this interval by some kind of clutch, implemented by means of two crown wheels (Figure 1; sectional view, *K1* and *K2*). As soon as the electrical circuit was closed, the electromagnets (Figure 1; rear view, *E*) clutched-in the crown wheels via a lever construction (Figure 1; sectional view, *m* and *H3*). When the voltage broke down, the crown wheels de-clutched and thus stopped the hands showing the elapsed time. The chronoscope was equipped with two dials, each of them divided into 100 scale units (Figure 1; frontal view). The hand of the upper dial made a full turn in a tenth of a second, while the hand of the lower dial made a full turn in 10 seconds. So, the upper dial indicated milliseconds, and the lower indicated tenths of a second. The escapement of this device was a spring (Figure 1; sectional view, *F*) oscillating at 1000 Hertz and so allowing the clockwork to make thousand steps per second.

In the 1860s, the astronomer Adolph Hirsch (1830–1901) was the first to use the chronoscope for non-technical time measurements in Neuchâtel (Switzerland). He used two instruments, manufactured by his friend Matthias Hipp. Hirsch measured the fastest possible reaction (the purely physiological reaction time) to auditory, visual, and tactile stimulation, and determined values of about 200 milliseconds for these three modalities (Hirsch 1865).

At almost the same time, the physiologist Frans Cornelis Donders (1818–1889) conducted experiments together with his doctoral student Johan Jacob de Jaager at Utrecht (the Netherlands) to determine the time expenditure for stimulus discrimination and reaction choice by using the so-called "noëmatachograph", literally a swiftness-of-thought writer (de Jaager 1865; Donders 1868). Experimenter and participant sat in front of a gramophone-like sound cone, to which a membrane was attached at its lower end, holding a needle to write the incoming focused sound waves to a band of sooted paper, wrapped around a rotating drum. The experimenter spoke the stimulus, and the participant reacted vocally according to the experimental task. The time elapsed between stimulus onset and reaction onset was gained in a cumbersome manner by counting the number of cycles a tuning fork had scratched into the paper band parallel to the vocal record. First, Donders and de Jaager measured the physiological reaction time to be about 200 milliseconds, and thus confirmed Hirsch's respective results. Then, they instructed the participant to react to one, and only one, stimulus from a randomly presented 5-element set of stimuli (namely the syllables *ka*, *ke*, *ki*, *ko*, *ku*) and in any other case to suppress the reaction. Finally, the participant had to react to every stimulus with its equivalent. Donders and de Jaager argued that the second task required stimulus discrimination in addition to the physiological reaction time, while the third task required discrimination effort *and* reaction choice effort in addition to the physiological reaction time. From these task-dependent reaction times, they calculated the time slices for stimulus discrimination and reaction choice at about 40 milliseconds according to their so-called "subtraction method".

More or less immediately after Wundt had established the Leipzig Institute in 1879, a young mathematician named Max Friedrich (1856–1887) began to measure reaction times to visual stimulation with the chronoscope (Friedrich 1883). As his doctoral thesis that emerged

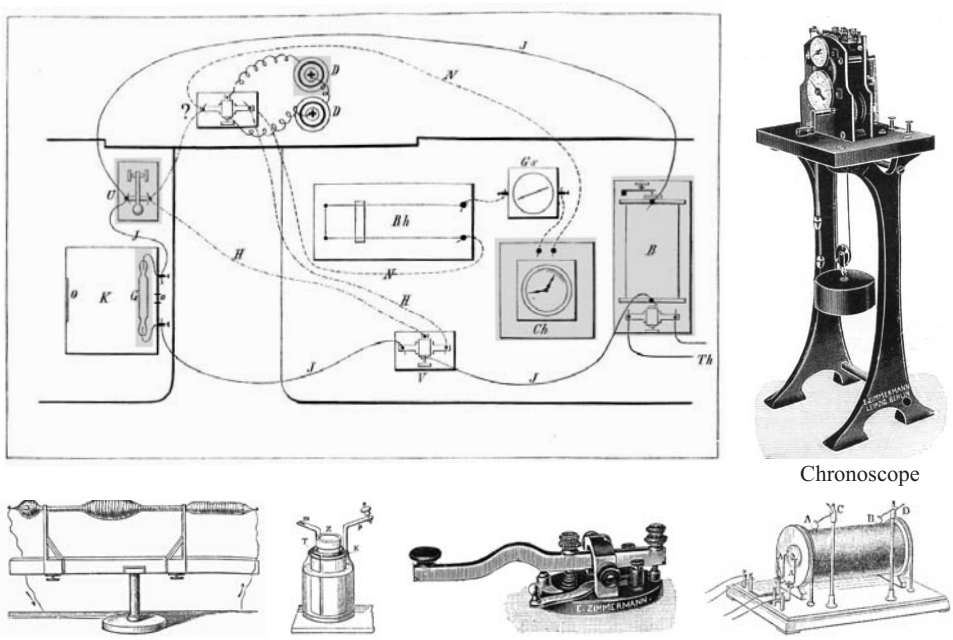


Figure 2. Friedrich's experimental setup shown in the upper left corner, as depicted in Friedrich's (1883) report. Moreover, one can see some important components in 3D-view, which constitute this setup. For a detailed explanation, please see the text. (Source Plan: Friedrich 1883: 45. Geißler tube: Meyer's 1893–97, 7th vol.: 240. Telegraph key: Zimmerman 1928: 123. Rühmkorff inductor: Meyer's 1893–97, 9th vol.: 224. Daniell element: Meyer's 1893–97, 7th vol.: 47. Chronoscope: see Figure 1.)

from these experiments is commonly seen as the very first dissertation in experimental psychology, Friedrich's setup will be explained in more details by means of the plan provided by Friedrich himself (see Figure 2): The visual stimulus was placed on the rear side *O* of a darkened box *K,* and the participant was instructed to look through a small hole *o* into this box. At the moment of presentation, the stimulus was lighted by a so-called Geißler tube[3] *G,* an ancestor of contemporary gas discharge tubes. A so-called Rühmkorff inductor[4] *R,* which transformed a low input voltage from a thermoelectric pile *Th* into a high output voltage, served as the high voltage source for the Geißler tube. Besides the high voltage circuit *J,* there was a low voltage circuit *N,* with a source in the form of two so-called Daniell elements[5] *D,* which are accumulators in today's terminology. The low voltage circuit supplied the chronoscope *Ch,* with electricity.[6]

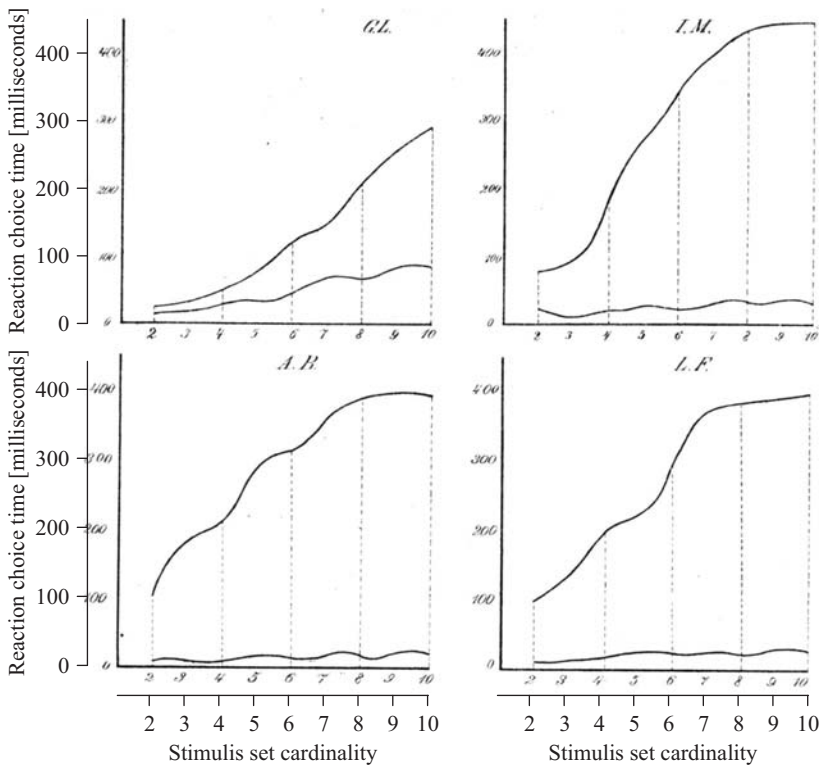
[3] Named after the German physicist and inventor Heinrich Geißler (1814–1879).

[4] Named after the German mechanic Heinrich Daniel Rühmkorff alias Ruhmkorff (1803–1877).

[5] Named after the English chemist John Frederic Daniell (1790–1845).

[6] Via the rheostat *Rh,* the amperage in this chronoscope circuit was adjustable to calibrate the time-measuring device. The respective amperage in this circuit could be seen on an amperemeter, *Gs*, which was called galvanoscope in the 19th century.

a)



b)

Figure 3. a) The so-called "psychophysical piano", constructed to take up the participant's stimulus-adequate reaction to a randomly presented stimulus from a set of stimuli. (Source: Photograph from the Wundt room at the University of Leipzig. © 2008 Maximilian Wontorra). *b)* Merkel's (1885) reaction choice times in milliseconds (ordinate) plotted against the stimulus set cardinality (abscissa) for four exemplarily selected participants. (Source: Merkel, 1885, Appendix, Table I, clipped from Figure 3. Axes re-labeled by © 2008 Maximilian Wontorra)

A typical trial was executed as follows: the participant was positioned in front of the box, and had to keep the telegraph key *U* pressed down. Now, the experimenter pressed the switch *V* to close the high voltage circuit, and to ignite the Geißler tube. Simultaneously, the chronoscope circuit was closed via the relay-like element, denoted by a question mark *?* (Friedrich did not explicitly explain this component), and the electric line *H.* Thus, stimulus presentation and the starting of the chronoscope were synchronized. The participant was instructed to react by releasing the key *U.* This action interrupted the high and low voltage circuit simultaneously. Thus, in the moment of releasing the key, the tube went out and the chronoscope stopped with the elapsed time between stimulus presentation and the participant's reaction.

With this setup, Friedrich again confirmed Hirsch's simple alias physiological reaction times at about 200 milliseconds. By utilizing Donders' and de Jaager's subtraction method, he determined the highly inter-individually as well as intra-individually varying recognition time for a color stimulus out of a 4-element set of stimuli in a range between 100 and 300 milliseconds. Furthermore, he found recognition times for 1-digit through 6-digit numbers, ranging from about 300 milliseconds for the 1-digit numbers to about 1.6 seconds for the 6-digit numbers, again highly varying between, as well as within, participants.

After the arrival of reaction time measurement at Wundt's institute, the field of application for chronometric investigations seemed nearly inexhaustible. Martin Friedrich Gottlob Trautscholdt (1883), for example, investigated the time expenditure of association processes, Ernst Tischer (1883) measured discrimination times for independently varied acoustic intensities, and Emil Kraepelin (1883a, b) investigated the influence of psychotropic drugs on reaction times.

Julius Merkel (1885) earned his doctorate at Wundt's institute in the mid-1880s with a study on reaction times to visual stimulation, also using a similar setup as Friedrich. Again, the exposition unit was a darkened box, in which the stimulus was exposed under instantaneous lighting. With this setup, Merkel determined, amongst others, the choice time of an adequate reaction to a stimulus out of a set of stimuli of cardinality 2 through 10. The stimuli were the first five Arabian and the first five Roman numbers. The participant was instructed to react to the 2 through 10 different stimuli with 2 through 10 different fingers via the so-called "psychophysical piano", an arrangement of two multiple telegraph keys (see Figure 3a).

Figure 3b shows Merkel's findings of four participants' reaction choice times plotted in milliseconds as a function of stimulus set cardinality (G. L. = Gustav Lorenz, I. [J.] M. = Julius Merkel himself [?], the others are not assignable). From these plots, one can see that choice times do not increase linearly with the number of available stimuli, but these times needed to identify a target stimulus are curvilinear-convex functions of the number of available stimuli. With this result, Merkel anticipated in some respects William Edmund Hick's (1912–1974) Law[7] by more than 60 years, at least qualitatively.

After six or seven years of reaction time studies at the Leipzig institute, this initially so promising research attempt ran into problems mainly in the course of the investigations by Wundt's doctoral students James McKeen Cattell and Gustav Oskar Berger, who determined time expenditures for discrimination operations close to, in some cases even *less than* (!) zero

[7] Hick's (1952) Law says that the search time for a target in a set of stimuli equals the dyadic logarithm of the cardinality of this respective set. This suggests an interpretation that we do not scan the set in question serially, but that we walk down a binary search tree, so to speak graph-theoretically.

(see Berger 1886; Cattell 1886a, b, c; 1888; see also Cattell 1886–87a, b, c; Wontorra 2008: 101–120). After these disquieting results, a theoretical reorientation took place. Leipzig researchers bit by bit abandoned Wundt's up-to-then almost dogmatically restated strictly serial approach of information processing in favor of a new view, thinking of mental operations as overlapping in time (see e.g., L. Lange 1888).

Attempts to quantify the phenomena of consciousness

As the concept of consciousness was one of the salient characteristics of Wundt's psychology, another important research line at Wundt's institute was concerned with the phenomena of consciousness and their quantification. According to Wundt's view, consciousness was the total content of our immediate experience (see e.g., Wundt 1908–11, 3rd vol.: 296). Wundt's consciousness consisted of an inner visual field, and an inner visual focus. The entry of a single idea from the field to the focus was named "apperception" in the terminology of the 19th century.[8] What Wundt and his contemporaries called apperception can be called today, more or less synonymously, attention, and what they called "apperceptual focus" comes closest to the concept of short-term, or working, memory in current terminology.

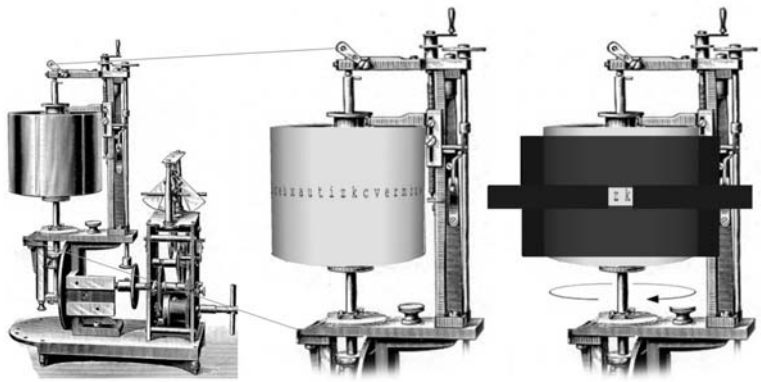
Working memory capacities

One important issue concerned the question of how many single ideas can be held simultaneously in the mentioned focus. The obvious method to answer this question was to increase a set of stimuli in respect to its cardinality, and expose this set tachistoscopically[9] as long as the participant was no longer able to replicate the single items from this set.

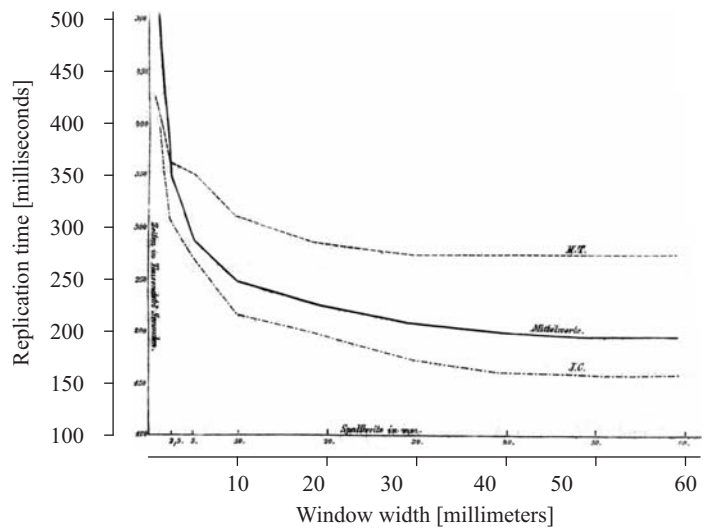
In 1885, the above-mentioned James McKeen Cattell (1860–1944) from Easton (PA), USA, published the first of a longer series of investigations in Wundt's journal *Philosophische Studien [Philosophical Studies].* Like other researchers in the realm of mental chronometry, in this and his following studies, which were combined in 1886 in a dissertation entitled *Psychometrische Untersuchungen [Psychometric investigations],* Cattell was primarily interested in the determination of the exact time expenditure of single mental operations. But in the 1885 study, entitled *Über die Zeit der Erkennung und Benennung von Schriftzeichen, Bildern und Farben [On the time needed for the recognition and naming of characters, pictures, and colors],* he also found, more or less incidentally, an only slightly inter-individually varying – in current terminology – buffer size of working memory. This investigation is worth mentioning for at least one more reason, namely, that it is hardly to be surpassed in the simplicity of its experimental setup, as well as in its rationale. In this respect, it contrasts pleasantly with some other investigations at Wundt's institute from the 1880s, where the technical effort was not always immediately paralleled by the complexity of the question to be answered. Cattell justi-

[8] Referring to a concept which can be traced back to Gottfried Wilhelm Leibniz (1646–1716).

[9] Tachistoscopes were devices designed for the short-term presentation of visual stimulus arrangements. In tachistoscopes, a blind with a window in it was moved in front of the mentioned arrangement. For the time of the window's passage, the arrangement was exposed short-term to the participant.



a)



b)

Figure 4. a) Cattell's (1885) experimental setup to determine the time taken by the recognition and vocalization (in short: replication) of characters. For a detailed explanation, please see the text. (Source: Zimmermann 1928: 185; post-processed by © 2009 Maximilian Wontorra). *b)* Plots of the replication time of a single character as a function of the number of simultaneously visible characters (4b). For a detailed explanation, please see the text. (Source: Cattell 1885: 638. Axes re-labeled by © 2009 Maximilian Wontorra)

fied his simple setup in the introduction to his study. He argued that the complex setups used so far had to be adjusted and maintained in a difficult and time-consuming manner, that the exposition units only rarely produced the stimulus to a satisfactory quality,[10] that the chrono-

[10] Needless to say, the sudden change of luminance, associated with the method of instantaneously lighting the stimulus in a darkened box, forced the beholder's eye to an adaptive task, taking time in its own right, so that the time taken by the mental act in question seemed to be longer than it was in fact.

scope did not measure times exactly enough,[11] and that the participants' task was artificial and thus far removed from everyday tasks, which would increase the risk that in some cases *not the total* time effort, and in other cases *more than* the time effort of the mental operation in question was determined (see Cattell 1885: 635). Surely, these methodological objections were justified. This makes it all the more astonishing that, during his series of investigations, Cattell's setups grew just as complex as those of his time-measuring predecessors. However, this matter is of no relevance to the study under discussion here.

To determine the time taken by the recognition and vocalization of characters (in short: the replication time), Cattell "misused" a recording instrument known as kymograph (see Figure 4a) for his experiments. As is well-known, the kymograph was a clockwork-driven drum, on which 19th-century experimental physiologists recorded the time series of physiological magnitudes such as respiration parameters, pulse rate, and so forth. The kymograph drum with a diameter of 50 centimeters was covered by Cattell with a band of white paper showing a random series of characters, as depicted above. Finally, Cattell placed a blind with a window, 1 centimeter in height and of variable width, in front of the drum. The height of the characters and the spacing were chosen so that for a width of x centimeters, x characters were visible simultaneously. If the width was less than 1 centimeter, only one character moved through the participant's visual field at a time. This was followed by a short pause, with the next character then appearing and vanishing again. With the participant positioned in front of the drum in such a way that he or she could comfortably see the single character or a subset of the total available characters in the window, Cattell slowly started the drum's rotation clockwise to increase the angle velocity of the drum. This kept being increased bit by bit, until the participant was overtaxed, and was no longer able to properly replicate the 30 or 40 characters moving through his or her visual field.

Cattell varied the window width independently, and determined for each particular width the respective limit velocity. From the window width and the limit velocity, he then calculated the (mean) time taken to replicate *one single* character as a function of window width. Cattell was vague in his explanation of this calculation, but it is to be assumed that he considered the replication time of a single character to be the time needed for moving forward the visible subset of letters by exactly one element.

Figure 4b shows plots of the replication times for the participants H. T. and J. C. (probably James McKeen Cattell himself) as a function of window width (dashed lines), and it shows the respective course of the means (solid line), averaged over the – in total – nine participants. One can see that the replication time decreases with window width, but beyond a width of 40 millimeters, i.e., four simultaneously visible characters, practically no further time gain is observable. In other investigations, not published until 1886, Cattell had determined the recognition time and the vocalization time for a single character in daylight conditions at about 250 and 100 milliseconds, respectively. By using a self-constructed device, which he called "Fallchronometer" [fall or gravity chronometer], he was able to stimulate tachistoscopically

[11] The Hipp-type chronoscope was a 19th-century state-of-the-art chronometer but it had to be calibrated very exactly. Moreover, it had to be checked rotationally during the experimental sessions, and it had to be protected against the permanent magnetization of the iron kernels of its coils by changing the direction of the direct current betimes. This was not absolutely guaranteed at least during the early years at Wundt's institute. Thus, the chronoscopically measured times of the early Leipzig investigations could not unconditionally be trusted (cf. Wontorra 2008: 65ff.).

and to measure exposition times. Therefore, Cattell argued, it is not a big surprise that replication times at a window width of 10 millimeters, i.e., exactly one visible character at a given moment, equaled fairly accurate 250 milliseconds. As in this case, both recognition and vocalization were completely automated processes, the respective previous character could be vocalized, while the respective next character was apperceived. In the case of two or more simultaneously visible characters, the process of vocalization and the several processes of apperception were overlapping up to the boundary value of about four simultaneously visible elements. According to Cattell's interpretation, with four elements the maximum of simultaneously processable impressions was reached. Thus, a numerical value was identified by Cattell as a real capacitive limitation of our mental machinery that corresponds astonishingly well to the *Magical number 4* of recent findings (cf. Cowan 2001, 2005) and which seems to be even a better estimate of the real limitation of short-term memory than George A. Miller's (1956) famous *Magical number 7, plus or minus 2*.

Awareness of ideas

To complicate things, in Wundt's psychology, the question of the number of momentarily present ideas was not an all-or-none problem, but each and every idea present at a given moment had a certain degree of awareness, or clearness.

Consequently, Wilhelm Wirth (1876–1952), the co-director of the Leipzig institute since 1908, designed an apparatus he called "Spiegeltachistoskop" [mirror tachistoscope] (see Figure 5, right part) at the turn of the century. This apparatus allowed a short-term modification of single elements in a permanent stimulus array, as illustrated in the left part of Figure 5, which shows the magnified element *O1* from the tachistoscope construction. By a slight modification of single elements, as, for example, the omission of the center of one of the depicted circles, and the detection probability for these changes by the participant in short-

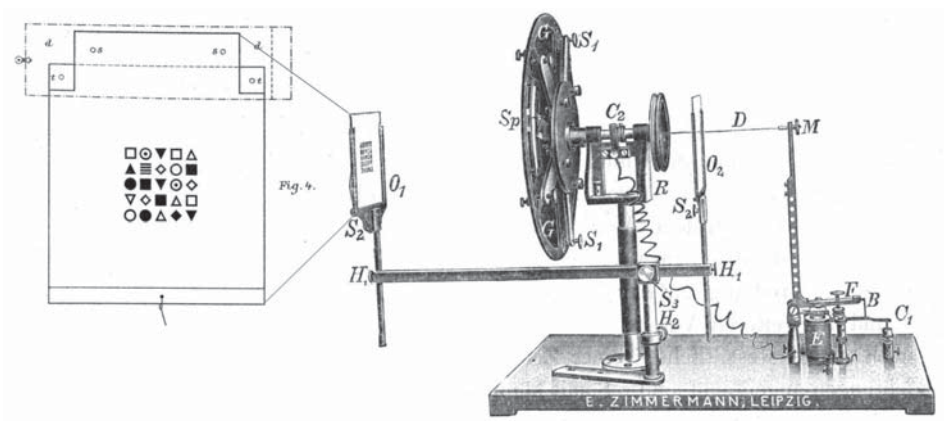
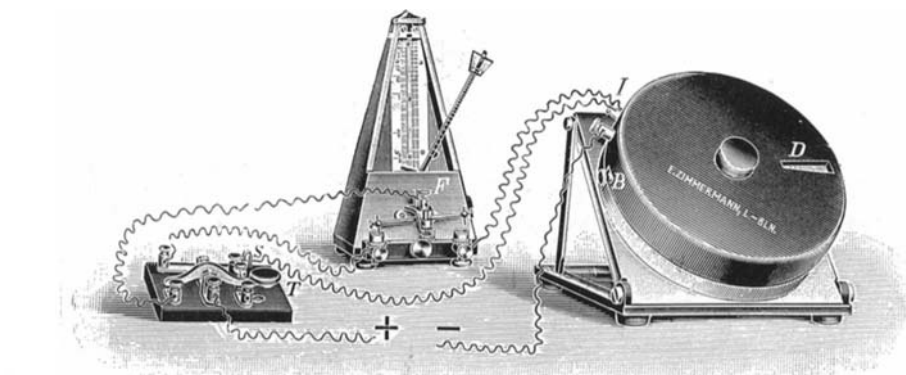


Figure 5. Wirth's mirror tachistoscope (right side) with a magnification of the stimulus array *O1* (left side). For a detailed explanation, please see the text. (Source: Wirth 1902, Appendix, Table II, Figure 4, and Wirth 1903: 688, respectively

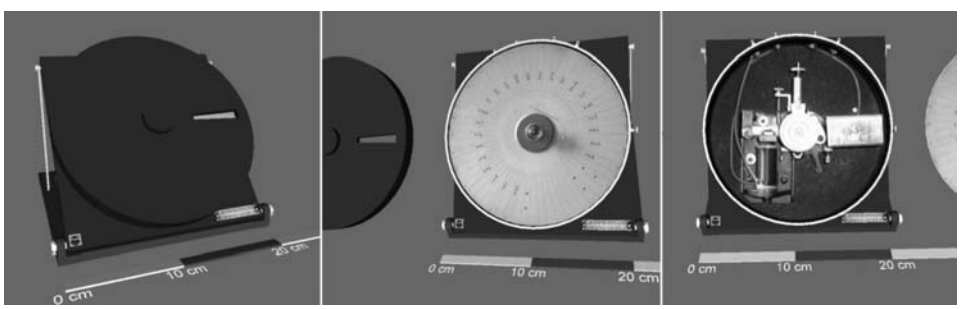
term presentation, Wirth (1902, 1903) wanted to determine the awareness of the respective element. The participant sat to the left of a rotating mirror *Sp,* looked into it, and thus permanently saw the mirror image of the original stimulus array *O1*. The experimenter sat on the right (behind the construction). By pulling cable *D,* the experimenter opened a window in the rotating mirror, so that the participant could have a short glance at the modification *O2* every time the gap in the mirror passed the participant's line of sight. Needless to say, the original and the modification had to be positioned in such a way that they stayed congruent over time except for the experimentally varied changes.

"Volatility" of the contents of consciousness

Of course, the number of momentarily present contents of consciousness as well as the degree of awareness of a single content are only snapshots. When Hermann Ebbinghaus (1850–1909)



a)



b)

Figure 6. a) A simple setup for the standardized presentation of stimuli in memory investigations. For a detailed explanation, please see the text. (Source: Zimmermann 1928: 64). *b)* Three screenshots of a virtualized Ranschburg-type mnemometer, which was widely used in the mentioned memory investigations. For detailed explanation, please see the text. (Source: Screenshots of a virtualized Ranschburg-type mnemometer. © 2003–2009 Maximilian Wontorra)

had published his seminal studies on memory contents and their time-dependent "dissolution processes"[12] in 1885, one had to be interested in the time characteristics of these volatile contents of consciousness (see Ebbinghaus 1885).

Figure 6a depicts a simple setup for the standardized presentation of item series in learning and retrieval experiments. The setup consists of a simple telegraph key, a metronome equipped with an interrupting mechanism, and a so-called "mnemometer" (literally a memory meter), constructed by Pál Ranschburg (1870–1945), who was a pioneer in Hungarian psychology. All these components are integrated into an electrical circuit. Reaching its maximum deflection, the oscillating metronome briefly closed the electrical circuit, and this impulse caused the stimulus change in the mnemometer (assuming the setup was activated by the experimenter via the telegraph key). Figure 6b consists of some screenshots of a virtualized Ranschburg-type mnemometer. The left shot shows the closed device. The center shot shows the stimulus card after the removal of the apparatus's lid, while the right shot allows a view into the device's innermost card-holding mechanism that consisted in a locked axis under spring tension, which was for the time of the electrical impulse (temporarily) unlocked by an electromagnet and thus enabled to rotate the stimulus card by one step to present the next stimulus in the lid's window.

Psychophysics of time perception

Based on Ernst Heinrich Weber's (1795–1878) investigations from the first half of the 19th century, resulting in Weber's constant ratio of, on the one hand, the stimulus *S,* and, on the other hand, the increment on this stimulus, necessary to produce a just noticeable perceptual difference, in short: $\frac{\Delta S}{S} = const.,$ Gustav Theodor Fechner (1801–1887) opened a new field of research with his *Elements of psychophysics* (1860). As is well-known, psychophysics is concerned with finding functions, mapping *physical* to *perceived* intensities. As Weber and Fechner were important predecessors of Wundt at the University of Leipzig in establishing empirical methods in psychology, it was in a sense straightforward for Wundt and his coworkers to conduct experiments in the realm of psychophysics, too. Mainly inspired by the experiments conducted by Ernst Mach (1838–1916) and Karl von Vierordt (1818–1884) (see Vierordt 1868), the experiments at Wundt's institute focused on the investigation of the "Zeitsinn" [time sense], as they called the psychophysics of time perception in late 19th century.

At Wundt's institute, researchers predominantly studied the psychophysics of time perception in the auditory modality. The first Leipzig investigation in this field was conducted by Julius Kollert (1883). Kollert who was convinced to have detected methodological weaknesses in Mach's und Vierordt's experiments, used two metronomes, like his criticized predecessors: the first metronome ticked a standard time interval, while the second ticked a target interval in the participant's back. These intervals were separated by a short time span. The length of

[12] It has to be noted that already before Ebbinghaus, the American physicist Francis E. Nipher (1847–1926) investigated memory processes, but he used numbers in contrast to Ebbinghaus, who experimented, as is commonly known, with senseless syllables (cf. Nipher 1876, 1878).

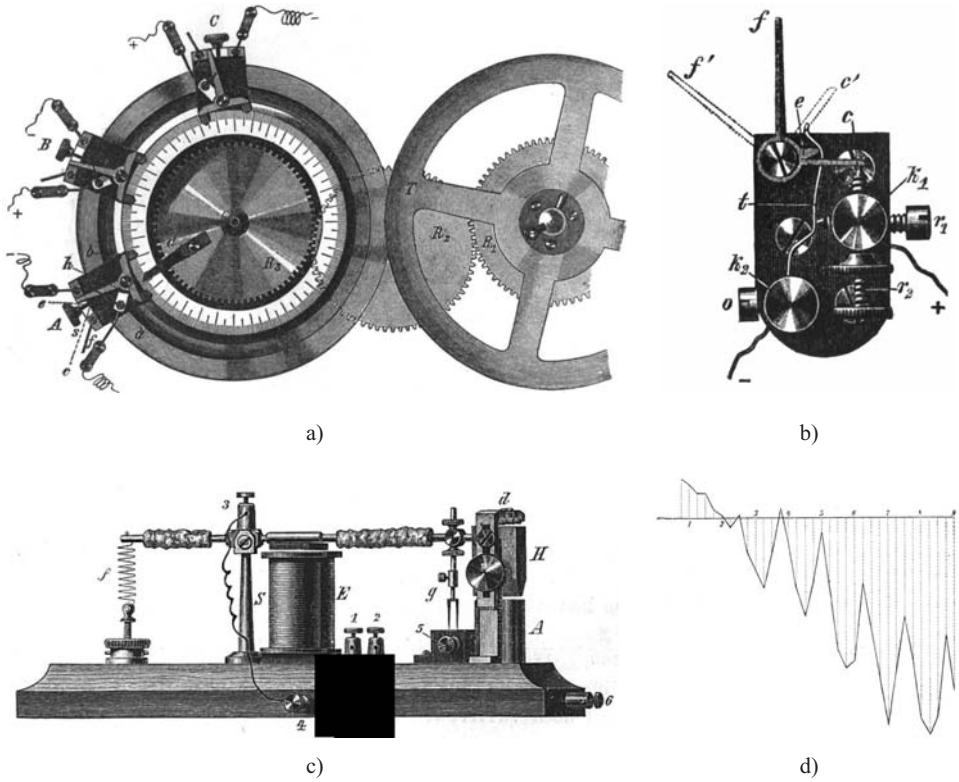


Figure 7. a) A so-called time sense apparatus. For a detailed explanation, please see the text. (Source: Wundt 1908–11, 3rd vol.: 344). *b)* A switch as it was attached to the static periphery of the time sense apparatus. For a detailed explanation, please see the text. (Source: Wundt 1908–11, 3rd vol.: 478). *c)* An electro-magnetically triggered sound hammer. For a detailed explanation, please see the text. (Source: Meumann 1894: 272). *d)* A function plot of Glass's (1888) so-called "constant error" (as the discrepancy between perceived and physical lengths of time intervals) against the physical length of a time interval. For a detailed explanation, please see the text. (Source: Glass, 1888: 454)

standard interval was varied independently, and, starting with a target interval equaling the length of standard interval, the experimenter varied the length of the target interval according to Fechner's so-called "method of minimal changes" as long as the participant perceived both interval lengths as equal. By pooling data over the participants, plotting the target-standard differences Δ against the standards *t,* and finally determining a best-fit curve according to the methods of least squares, Kollert found this exponential psychophysical function . Here, *e* is the Euler number, and *a, b* are parameters, fitted by Kollert to his sample as $a = 0.1021$, and $b = 0.0480$. With these parameters, Kollert's connection between the "estimation error", Δ (as it was initially called by Wundt and his colleagues) and the physical length of an interval *t,* was a smooth, monotonically decreasing function. Short physical intervals were subjectively overestimated, and long intervals were underestimated, with a point of indifference at a physical length of about 0.75 seconds. Kollert's exponential function was at least implausible in

respect to the fact that with growing physical lengths, the discrepancies between the objective lengths and the perceived ones would rapidly approach negative infinity.

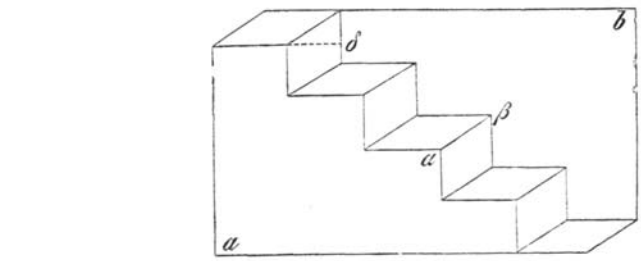
Mainly in order not to be restricted to relatively short intervals realizable by means of metronomes, Kollert's successors in the realm of time sense investigations experimented with a new stimulus generator, which they called "Zeitsinnapparat" [time sense apparatus], as depicted in Figure 7a. This Meumann-type[13] apparatus consisted in a gear-driven rotating disc, to which a pin was attached. As the disc rotated, the pin touched switches, similar to that depicted in Figure 7b. These switches were mounted onto the static peripheral ring of the time sense apparatus. As soon as one of these switches was touched by the pin, an electrical circuit was closed, which in turn triggered a beat of an electro-magnetic sound hammer, as it can be seen in Figure 7c. With three of the mentioned switches, two intervals were realizable. In this case, the center switch and the corresponding sound hammer beat marked the end of the standard and the beginning of the target interval, respectively. By means of this apparatus, several investigations were conducted at Wundt's institute, each one criticizing the previous investigation in respect to its methodological weaknesses. In 1887, Paul Richard Glass submitted his time sense investigation at Wundt's institute as a doctoral thesis (cf. Glass 1888). Figure 7d shows his function, again plotting the now misleadingly so-called "constant error", as the target-standard difference (ordinate) against physical time spans in seconds (abscissa), with a y-axis, ten times overscaled in comparison to the x-axis. Glass's psychophysical function exhibits a linearly decreasing trend, and a superimposed oscillatory component with a frequency of about 1 Hz. Despite all divergences in detail, all of Kollert's successors found this linear trend with a point of indifference (the function's root) between 2 and 3 seconds, which is close to the value found in current investigations on the duration of the psychological present (see for instance Pöppel 2004). Moreover, all researchers found this peculiar oscillatory component.

Apperceptual waves

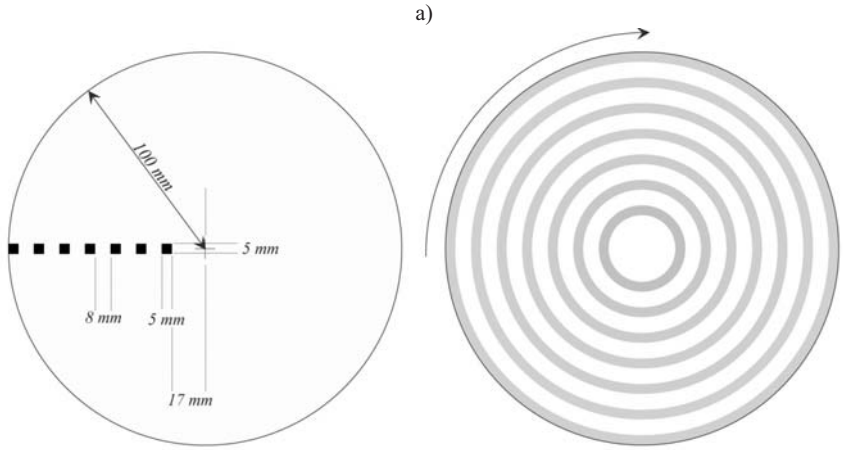
Mainly inspired by investigations conducted by the otologist Viktor Urbantschitsch (1847–1921), in Vienna, Austria, in the realm of auditory perception, in the second half of the 1880s, investigators at the Leipzig institute started to develop an interest in a phenomenon that can be traced back to David Hume (1711–1776), and his *Treatise of Human Nature* (1740). Wundt and his colleagues called this phenomenon "Apperzeptionswellen" [apperceptual or attentional waves]. It consisted in the fact that a stimulus (of low intensity) intermittently fades out to be re-perceived clearly after a certain time span. Hopefully, to have a post-hoc explanation for the periodicities in the above-mentioned psychophysical functions of time perception, Leipzig researchers started to investigate this phenomenon in the auditory, the visual, as well as the tactile modality.

The first extensive Leipzig investigation of the apperceptual waves was conducted by Nicolai Lange from St. Petersburg (Russia). N. Lange (1888) used tactile, auditory, and visual stimuli to determine primarily the chronoscopically measured time spans between the

[13] Named after Ernst Meumann (1862–1915), one of Wundt's assistants and life-long friends, who essentially enhanced previous models of this apparatus. For Meumann's own investigations, see Meumann (1893, 1894).



a)



b)

Figures 8. a) The Schröder staircase as an ambiguous figure, as used by N. Lange (1888). For a detailed explanation, please see the text. (Source: N. Lange 1888: 406). *b)* Masson's disc, as described and used by Pace (1893). For a detailed explanation, please see the text. (Source: Drawing of Masson's disc, as described and used by Pace (1893). © 2008–2009 Maximilian Wontorra

moments of maximal perceived clearness of the respective stimulus (the participant had to start or stop the chronoscope by pressing or releasing a telegraph key, respectively.) For the visual modality Lange used, among other stimuli, an ambiguous figure, called "Schrödersche Treppe" [Schröder staircase],[14] as depicted in Figure 8a. (The Schröder staircase is, like the "Necker cube", "Rubin's vase", or the "Spinning dancer",[15] is one of a multitude of ambiguous figures, giving rise to two different perceptions.) In the case of Schröder's figure, one either sees a staircase from underneath (concave), or a staircase from above (convex), and, upon continuous viewing, the impression alternates. Using the Schröder staircase, N. Lange determined the time spans during which the convex, or the concave version of the staircase was visible.

The general Leipzig view, according to which the reason for the apperceptual waves was primarily an attentional one, was not undisputed. Amongst others, for Wundt's former doc-

[14] Named after Heinrich Georg Friedrich Schröder (1810–1885), who described this figure in 1858 for the first time.
[15] See, for example, http://en.wikipedia.org/wiki/The_Spinning_Dancer. (Accessed: 29 November 2009).

toral student, and later on professor at Harvard, Hugo Münsterberg (1863–1916), as well as Alfred Lehmann (1858–1921), who was for a research stay at Leipzig in the mid-1880s, these waves were quite simply due to a temporary fatigue of the respective sense organ.

After having determined the periodicities of each of the diverse modalities separately, N. Lange tried to conduct, so to speak, an *experimentum crucis* to decide whether the central–attentional or the peripheral–adaptive explanation for the apperceptual waves was appropriate. To accomplish this, N. Lange stimulated the visual and the auditory channels simultaneously. The participant was instructed to press two buttons, one button to signal the moments of maximal clearness on the first, and the other to signal the respective moment on the second sensory channel. In this case, the participant's reactions were recorded kymographically. Finding roughly the same periodicities of fluctuation in this combined experiment as he had found in the previous single-modality investigations, this was sufficient evidence for N. Lange to express his conviction that the phenomenon of apperceptual waves was solely due to the periodical central shifts of attention.

To gather additional evidence against the peripheral explanation, another former doctoral student at the Leipzig institute, Edward A. Pace (1861–1938), conducted experiments with a device named "Masson's disc",[16] as shown in Figure 8b (Pace 1892). While rotating this disc with a sufficiently high velocity, the beholder sees a number of grey rings, decreasing in luminance with the distance to the disc's center. Arbitrarily assigning a luminance of value of 1 to the background of Masson's disc, Helmholtz (1867) calculated the luminance h of a specific ring according to the formula , with π being Ludolph's number, r being the disc's radius, and d being the distance between the center of the respective ring-generating black patch and the midpoint of Masson's disc. The participant's task was to gaze at one of the outer rings, and to report the perceptual fluctuations of the ring which was gazed at. In a subseries of experiments, Pace worked with participants whose eyes were atropinized. As even these participants reported the respective fluctuations, Pace argued that the putative temporary fatigue of the eye's adaptive system of muscles was not a sufficient explanation of the phenomenon.

To this day, researchers do not exactly know how to explain the bi-stable percepts associated with ambiguous figures (cf. Kornmeier 2007). Utilizing event-related potentials as for instance EEG records, they try to find out whether bottom-up (mainly physiological), or top-down (mainly psychological) processes are responsible for these percept changes. Most recently, the tiny oscillations of the eyes' axes, the so-called microsaccades, which have been considered as pure noise of the visual system for a long time, have returned to the focus of interest as indicators of an impending shift of attention. In eye-tracker studies, stimulating amongst others by means of "Troxler's figure",[17] researchers found that the frequency of microsaccades increased some milliseconds prior to the re-appearance of the stimulus (cf. Engbert and Kliegl 2003; Hafed and Clark 2002; Martinez-Conde et al. 2004, 2006; Martinez-Conde and Macknik 2007).

[16] Named after Antoine Philibert Masson (1806–1858/1860).

[17] Named after Ignaz Paul Vitalis Troxler (1780–1866). This figure consists of a light grey ring with a fixation cross in its center. Gazing at the cross, the ring intermittently disappears.

Processing multimodal stimuli

Investigations in the characteristics of processing disparate sensations were, in the end, just as the research line of reaction time studies, inspired by the astronomical problems mentioned at the beginning of this article. Wundt was interested in this topic from very early on, and he published an article already in 1862 in the family magazine *Gartenlaube [Arbor].* In this article, he introduced a method for the determination of the "swiftest thought" by means of a slightly modified standard pendulum clock (Wundt 1862). For this reason, he attached two clappers to the pendulum rod, each of them hitting a bell at the moment when the pendulum reached its maximum deflection to the left or to the right, respectively. Comparing the *de facto* position of the clock's second hand to the *perceived* position of the bell tone at that moment, he found a divergence of about eight tenth of a second. He interpreted this elapsed time to be his personal swiftest thought, for Wundt was convinced that the processing of the auditory stimulus distracted the attention from the visual impression for the mentioned time span. An enhancement of this rough-and-ready device was Wundt's "Complicationspendel" [complication pendulum], as depicted in Figure 9. In principle, this apparatus consisted in a hand *Z,* which was moved by a pendulum *P* in front of a semicircular scale *S.* By shifting the weight *L* along the rod *P,* the pendulum's period was adjustable in a range between 0.5 and 1.0 Hz. At any point on the hand's trajectory, a distracting auditory stimulus could be triggered

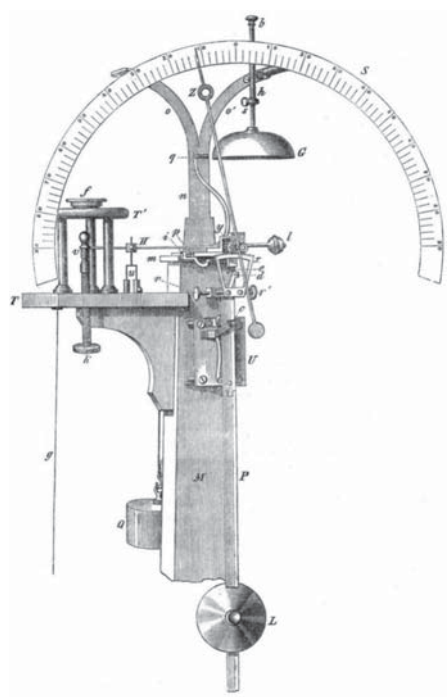


Figure 9. Wundt's "Complicationspendel" [complication pendulum]. For a detailed explanation, please see the text. (Source: Wundt 1908–1911, 3rd vol.: 72)

by beating the clapper *q* against the bell *G.* By only slightly modifying the apparatus, a tactile distractor in the form of an electric shock was applicable, too.

The first to conduct an extensive investigation on the processing of disparate stimuli (or "complications", as the late 19th century called multimodal processing) by means of Wundt's pendulum, was Woldemar von Tchisch, a physician from St. Petersburg in Russia, who experimented at the Leipzig institute in the 1880s (von Tchisch 1885). Von Tchisch varied independently the combinations of tactile, and/or auditory distractors, as well as the hand's position at which the distracting stimuli were triggered. With himself being his one and only participant in his experiments, von Tchisch produced huge data sets. One of the results gained from the data was that the greater the distractor-caused time shifts in perceiving the hand's position was, the lesser the pendulum's (and thus the hand's) velocity was. This is a highly contra-intuitive result. Assuming that it takes a certain amount of time to move one's attentional focus from the (visual) main sensation to the distracting sensation and back again to the main sensation (the central idea behind all these investigations concerned with the figuring out of the time characteristics of attentional shifts), one would expect that the smaller the divergence between the hand's de facto position and the perceived position was, the smaller the moving object's velocity was during this "blackout phase" of attentional shifts. Similarly to this result, nearly all of von Tchisch's results massively lacked credibility. This may be due to the constructional weaknesses of the stimulus generator. To circumvent these problems, Leipzig researchers developed variations of Wundt's original pendulum (Weyer 1898, 1900), or they experimented with a new clock-like construction called the complication clock (Geiger 1903). In this latter device, a weight-driven hand rotated at constant angle velocity in front of a clock face.

The last to experiment with Wundt's original pendulum was Christof D. Pflaum at the end of the 19th century (Pflaum 1900). Caroline Augusta Foley Rhys Davids (1857–1942) wrote a review of Pflaum's experiment in *Mind* (Rhys Davids 1899), saying, in essence, that Pflaum forgot to report a serious problem associated with this device, namely, that the hand of the device would always jump at the moment of triggering the distractor. Thereupon, Wundt (1900) wrote a huffish reply, accusing all those who unsuccessfully had tried to replicate the (mistrusted) Leipzig results by means of his pendulum to be incapable of handling this apparatus properly.

Summary

As the author hopes, it can be seen from this short sketch that researchers in early apparatus-based experimental psychology investigated a series of topics that are still relevant for today's psychology. After a long developmental history of methods and technology, contemporary researchers go for new results with enhanced methodological repertory, and updated devices, of course.

Perhaps the most important result regarding early reaction time studies was the fact that, after a couple of years, Leipzig experimenters were convinced that they had gathered sufficient evidence against Wundt's approach of strictly serial information processing. This brought into existence a new view which is similar to contemporary theories of information processing according to which particular mental operations take place simultaneously, and

are therefore undersummative in respect of the total amount of time individually consumed by each of the operations.

Within the group of consciousness studies, the investigations on the processing of multimodal stimuli did not produce any reliable results. This was mainly due to the exposition units for the visual main stimulus and the distracting tactile/auditory stimuli, as these units suffered from constructional defects that could not be overcome in Wundt's era. Leipzig researchers did not succeed in finding the causes of the so-called apperceptual waves. This is not at all surprising, as even contemporary researchers do not exactly know how to explain the perceptual fluctuations associated with ambiguous stimuli. Despite all the differences in individual function plots, in the realm of the psychophysics of time perception, Wundt and his colleagues consistently found that the lengths of short time spans were subjectively overestimated, while long spans were underestimated. Moreover, in these experiments they found a point of indifference which was close to the value obtained in current investigations on the duration of the psychological present. All in all, the early attempts to determine working memory limitations were the most convincing ones. Cattell, using a really simple setup, found values replicable in essence until today already in the mid-1880s.

References

Berger, G. O. (1886) Über den Einfluss der Reizstärke auf die Dauer einfacher psychischer Vorgänge mit besonderer Rücksicht auf Lichtreize. [On the influence of stimulus intensity on the duration of simple psychic processes with specific respect to light stimuli.] *Philosophische Studien [Philosophical Studies]* 3, 38–93.

Bessel, F. W. (1822) Astronomische Beobachtungen auf der Königlichen Universitätssternwarte in Königsberg. [Astronomical observations at the royal university's observatory at Königsberg.] Königsberg: Universitätsbuchhandlung.

Cattell, J. McKeen (1885) Über die Zeit der Erkennung und Benennung von Schriftzeichen, Bildern und Farben. [On the time of recognition and naming of characters, pictures, and colors.] *Philosophische Studien [Philosophical Studies]* 2, 635–650.

Cattell, J. McKeen. (1886a). Über die Trägheit der Netzhaut und des Sehzentrums. [On the response time of the retina and the center of vision.] *Philosophische Studien [Philosophical Studies]* 3, 94–127.

Cattell, J. McKeen (1886b). Psychometrische Untersuchungen. Erste Abteilung. [Psychometric investigations. First section.] *Philosophische Studien [Philosophical Studies]* 3, 305–335.

Cattell, J. McKeen (1886c). Psychometrische Untersuchungen. Zweite Abteilung. [Psychometric investigations. Second section.] *Philosophische Studien [Philosophical Studies]* 3, 452–492.

Cattell, J. McKeen (1886–87a). The time taken up by cerebral operations. *Mind* 11, 220–242.

Cattell, J. McKeen (1886–87b). The time taken up by cerebral operations. *Mind* 11, 377–392, 524–538.

Cattell, J. McKeen (1888) Psychometrische Untersuchungen. Dritte Abteilung. [Psychometric investigations. Third section.] *Philosophische Studien [Philosophical Studies]* 4, 241–250.

Cowan, N. (2001) The magical number 4 in short-term memory. A reconsideration of mental storage capacity. *Behavioral and Brain Sciences* 24, 87–185. [http://web.missouri.edu/~cowann/docs/articles/2001/Cowan%20BBS%202001.pdf]

Cowan, N. (2005) *Working Memory Capacity.* Hove, East Sussex, UK: Psychology Press.

Donders, F. C. (1868) Die Schnelligkeit psychischer Prozesse: Erster Artikel. [The velocity of psychic processes: First article.] *Archiv für Anatomie, Physiologie und wissenschaftliche Medizin [Archive for Anatomy, Physiology, and Scientific Medicine]* 657–681.

Ebbinghaus, H. (1885) Über das Gedächtnis. Untersuchungen zur experimentellen Psychologie. [On Memory. Investigations on Experimental Psychology.] Leipzig: Duncker & Humblot.

Engbert, R., and Kliegl, R. (2003) Microsaccades uncover the orientation of covert attention. *Vision Research* 43, 1035–1045.

Fechner, G. Th. (1860) Elemente der Psychophysik. Erster und zweiter Teil. [Elements of Psychophysics. First and Second Part.] Leipzig: Breitkopf & Härtel.

Friedrich, M. (1883) Über die Apperceptionsdauer bei einfachen und zusammengesetzten Vorstellungen. [On the duration of apperception for simple and complex ideas.] *Philosophische Studien [Philosophical Studies]* 1, 38–77.

Geiger, M. (1903) Neue Complicationsversuche. [New complication experiments.] *Philosophische Studien [Philosophical Studies]* 18, 347–436.

Glass, R. (1888) Kritisches und Experimentelles über den Zeitsinn. [Critique and experimental results concerning the time sense.] *Philosophische Studien [Philosophical Studies]* 4, 423–456.

Hafed, Z. M., and Clark, J. J. (2002) Microsaccades as an overt measure of covert attention shifts. *Vision Research* 42, 2533–2545.

Helmholtz, H. v. (1850a). Über die Fortpflanzungsgeschwindigkeit der Nervenreizung. [On the velocity of propagation of nervous stimulation.] *Annalen der Physik und Chemie [Annals of physics and chemistry]* 79, 329–330.

Helmholtz, H. v. (1850b). Messungen über den zeitlichen Verlauf der Zuckung animalischer Muskeln und die Fortpflanzungsgeschwindigkeit der Reizung in den Nerven. [Measurements on the time course of the contraction of animal muscles and on the velocity of propagation of nervous stimulation.] *Archiv für Anatomie, Physiologie und wissenschaftliche Medicin [Archive for Anatomy, Physiology, and Scientific Medicine],* 276–364.

Helmholtz, H. v. (1852) Messungen über Fortpflanzungsgeschwindigkeit der Reizung in den Nerven. [Measurements on the velocity of propagation of nervous stimulation.] *Archiv für Anatomie, Physiologie und wissenschaftliche Medicin [Archive for Anatomy, Physiology, and Scientific Medicine],* 199–216.

Helmholtz, H. v. (1867) *Handbuch der physiologischen Optik* [Handbook of Physiological Optics.] Leipzig: Voss.

Hick, W. E. (1952) On the rate of gain of information. *Quarterly Journal of Experimental Psychology* 4, 11–26.

Hirsch, A. (1865) Chronoskopische Versuche über die Geschwindigkeit der verschiedenen Sinneseindrücke und der Nerven-Leitung. [Chronoscopical experiments on the velocity of different sensations and on nervous propagation.] *Untersuchungen zur Naturlehre des Menschen und der Thiere [Investigations Concerning the Physics of Man and Animal]* 9, 183–199.

Hume, D. (1740) *A Treatise of Human Nature.* [http://socserv2.socsci.mcmaster.ca/~econ/ugcm/3ll3/hume/treat.html]

de Jaager, J. J. (1865) De physiologische tijd bij psychische processen. [The Physiological Time Taken by Psychic Processes.] Utrecht: P. W. van de Weijer.

Kollert, J. (1883) Untersuchungen über den Zeitsinn. [Time sense investigations.] *Philosophische Studien [Philosophical Studies]* 1, 78–89.

Kornmeier, J. (2007) *Ambiguous Figures: Evidence for Weak Neural Representation and Two Independent Processes.* Talk at the research colloquium, hosted by the Psychological Departments of the University

of Leipzig, 12 November 2007. [Vortrag im Rahmen des Forschungskolloquiums der psychologischen Institute der Universität Leipzig am 12. 11. 2007.]

Kraepelin, E. (1883a). Über die Einwirkung einiger medicamentöser Stoffe auf die Dauer einfacher psychischer Vorgänge. Erste Abteilung. Über die Einwirkung von Amylnitrit, Aethyläther und Chloroform. [On the influence of some medicamentous substances on the duration of simple psychic processes. First section: On the influence of amyl nitrite, ethyl ether, and chloroform.] *Philosophische Studien [Philosophical Studies]* 1, 417–462.

Kraepelin, E. (1883b). Über die Einwirkung einiger medicamentöser Stoffe auf die Dauer einfacher psychischer Vorgänge. Zweite Abteilung. Über die Einwirkung von Aethylalkohol. [On the influence of some medicamentous substances on the duration of simple psychic processes. Second section: On the influence of ethyl alcohol.] *Philosophische Studien [Philosophical Studies]* 1, 573–605.

Lange, L. (1888). Neue Experimente über den Vorgang der einfachen Reaction auf Sinneseindrücke. [New experiments on the process of simple reaction to sensation.] *Philosophische Studien [Philosophical Studies]* 4, 479–510.

Lange, N. (1888). Beiträge zur Theorie der sinnlichen Aufmerksamkeit und der activen Apperception. [Contributions to a theory of sensory attention and active apperception.] *Philosophische Studien [Philosophical Studies]* 4, 390–422.

Martinez-Conde, S., and Macknik, S. L. (2007) Windows on the mind. *Scientific American* 297, 56–63.

Martinez-Conde, S., Macknik, S. L., and Hubel, D. H. (2004) The role of fixational eye movements in visual perception. *Nature Reviews Neuroscience* 5, 229–240.

Martinez-Conde, S., Macknik, S. L., Troncoso, X. G., and Dyar, T. A. (2006) Microsaccades counteract visual fading during fixation. *Neuron* 49, 297–305.

Merkel, J. (1885) Die zeitlichen Verhältnisse der Willensthätigkeit. [The temporal relations of arbitrary activity.] *Philosophische Studien [Philosophical Studies]* 2, 73–127.

Meumann, E. (1893) Beiträge zur Psychologie des Zeitsinns. [Contributions to a psychology of time sense.] *Philosophische Studien [Philosophical Studies]* 8, 431–509.

Meumann, E. (1894) Beiträge zur Psychologie des Zeitsinns. (Fortsetzung.) Zweiter Abschnitt. [Contributions to a psychology of time sense. (Continuation.) Second section.] *Philosophische Studien [Philosophical Studies]* 9, 264–306.

Meyers (1893–97) *Konversationslexikon*. 5. Auflage. 17 Bde. Leipzig und Wien: Bibliographisches Institut. [Meyer's (1893–97) *Conversational Encyclopedia*. 5th edition. 17 vols. Leipzig & Vienna: Bibliographical Institute.]

Miller, G. A. (1956) The magical number seven, plus or minus two. Some limits on our capacity for processing information. *Psychological Review* 63, 81–97. [http://www.musanim.com/miller1956/]Nipher, F. E. (1876) Probability of error in writing a series of numbers. *American Journal of Science and Arts* (3rd series) 12, 79–80.

Nipher, F. E. (1878) On the distribution of errors in numbers written from memory. *Transactions of the Academy of Science of St. Louis* 3, 110–111.

Pace, E. (1893) Zur Frage der Schwankungen der Aufmerksamkeit nach Versuchen mit der Masson'schen Scheibe. [On the topic of attentional fluctuations after experiments by means of Masson's disc.] *Philosophische Studien [Philosophical Studies]* 8, 388–402.

Pflaum, C. D. (1900) Neue Untersuchungen über die Zeitverhältnisse der Apperception einfacher Sinneseindrücke am Complicationspendel. [New investigations on the time characteristics of the apperception of simple sensations by means of the complication pendulum.] *Philosophische Studien [Philosophical Studies]* 15, 139–148.

Pöppel, E. (2004) Lost in time. A historical frame, elementary processing units and the 3-second window. *Acta Neurobiologiae Experimentalis* 64, 295–301.

Rhys Davids, C. A. F. (1899). Philosophische Studien. [Philosophical Studies.] *Mind* 8: 4, 563–564.

von Tchisch, W. (1885) Über die Zeitverhältnisse der Apperception einfacher und zusammengesetzter Vorstellungen, untersucht mit Hülfe der Complicationsmethode. [On the time course of apperception in respect to simple and complex ideas, investigated by means of the complication method.] *Philosophische Studien [Philosophical Studies]* 2, 603–634.

Tischer, E. (1883) Über die Unterscheidung von Schallstärken. [On the discrimination of sound intensities.] *Philosophische Studien [Philosophical Studies]* 1, 495–542.

Trautscholdt, M. (1883) Experimentelle Untersuchung über die Association der Vorstellungen. [Experimental investigation on the association of ideas.] *Philosophische Studien [Philosophical Studies]* 1, 213–250.

Vierordt, K. v. (1868) Der Zeitsinn nach Versuchen. [The Time Sense after Experiments.] Tübingen: Laupp.

Weyer, E. Moffat. (1898) Die Zeitschwellen gleichartiger und disparater Sinneseindrücke. [The time thresholds of similar and disparate sensations.] *Philosophische Studien [Philosophical Studies]* 14, 616–639.

Weyer, E. Moffat. (1900) Die Zeitschwellen gleichartiger und disparater Sinneseindrücke (Schluss). [The time thresholds of similar and disparate sensations (End).] *Philosophische Studien [Philosophical Studies]* 15, 67–138.

Wirth, W. (1902) Zur Theorie des Bewusstseinsumfanges und seiner Messung. [On the theory of the scope of consciousness and its measurement.] *Philosophische Studien [Philosophical Studies]* 20, 487–669.

Wirth, W. (1903) Das Spiegeltachistoskop. [The mirror tachistoscope.] *Philosophische Studien [Philosophical Studies]* 18, 687–700.

Wontorra, M. (2008) Fragestellungen und Versuchsaufbauten der frühen apparativen Psychologie. Eine methoden- und apparateanalytische Untersuchung zum Forschungsprogramm an Wundts Leipziger Institut. [Scientific objectives and experimental setups in early apparatus-based psychology. A method- and apparatus-analytical investigation of the research program at Wundt's Leipzig institute.] Leipzig: Dissertation.

Wundt, W. (1862) Die Geschwindigkeit des Gedankens. [The swiftness of thought.] *Gartenlaube [Arbor]* 263–265.

Wundt, W. (1900) Zur Technik des Complicationspendels. [On the complication pendulum's technique.] *Philosophische Studien [Philosophical Studies]* 15, 579–582.

Wundt, W. (1908–11) *Grundzüge der physiologischen Psychologie. [Principles of Physiological Psychology.]* 6., umgearbeitete Auflage. 3 Bde. [6th, revised edition. 3 vols.] Leipzig: Engelmann.

Zimmermann, E. (1928) *Wissenschaftliche Apparate*. Liste 50. [Scientific Apparatuses. List No. 50.] Leipzig: Heine.

INTERDISCIPLINARY ISSUES IN EARLY CYBERNETICS

Leone Montagnini

Introduction[1]

What was the role played by interdisciplinarity in early cybernetics? By the so called *early cybernetics* we mean the cybernetics from its origins, during World War II until the mid-1950s. Before that *second cybernetics, artificial intelligence,* and much later *cognitive science,* would establish themselves as new successful paradigms in approximately the same territories of facts. One could rightly find the question interesting as a case study because of the importance that interdisciplinarity maintains in the other paradigms, and also in the present situation of science. Nonetheless, the question involves a very specific concern too, that is, whether interdisciplinarity did not represent just the very Achilles' heel of cybernetics, in this way deciding its fate. Such a question was raised especially in Italy during the 1970s, as de Luca (2006) reminds us.

We shall begin by focusing on interdisciplinarity in Norbert Wiener's works as a representative of the way of thinking of the main "early cyberneticians", in particular of Warren McCulloch, Walter Pitts, John von Neumann, remarking the divergences wherever possible. After that, we shall review the problems that emerged during the Macy Conferences on Cybernetics, as far as interdisciplinarity is concerned.

The general stance here is that an interdisciplinary approach was essential to cybernetics in itself, as a program in which different scientific traditions were fruitfully integrated. Nonetheless, the classical problems arising in the interdisciplinary collaboration afflicted it from the outset. In any case, it seems that these difficulties were not enough to justify the abandonment of the program.

Individual interdisciplinarity

Wiener (1950) claimed: "We need a range of thought that will really unite the different sciences, shared among a group of men who are thoroughly trained, each in his own field, but who also possess a competent knowledge of adjoining field" (57). These statements

[1] Acknowledgement: I gratefully acknowledge for the useful discussions about the themes treated here with the members of the School of Cybernetics founded in Naples by Eduardo R. Caianiello, in particular Settimo Termini, Aldo De Luca, and Giacomo Della Riccia, who was also Norbert Wiener's last collaborator at MIT. These same feelings I should like to address to the late Antonio Lepschy, one of the two engineers who introduced the theory of automatic control in Italy.

synthesize well Norbert Wiener's ideas on interdisciplinarity, regarding two interweaving levels: the interdisciplinarity in the individual, and the interdisciplinarity in the group.

Talking about the first, Wiener's requirements were made for a kind of scientist who was enabled by a wide educational background to interact with scientists of other disciplines and nevertheless possessed a clear disciplinary identity. With regard to the second, we would need to distinguish between very small work groups which carried out specific projects on experimental and mathematical research, and larger interdisciplinary groups which would have been considered useful for discussions and information interchange.

These ideas had very deep autobiographical roots. Personally, Wiener had wide cultural and scientific bases combined with deep roots in a particular field, i.e., mathematics. He had received a B.S. in mathematics and a Ph.D. in philosophy, studying with philosophers and scientists considered among the best minds in the very beginning of the 20th century. Afterwards, he joined the MIT Department of Mathematics, where he stayed for over 40 years, and entitled the second volume of his autobiography as "I am a Mathematician" (Wiener 1964).

Wiener's interdisciplinary personal formula was also typical of many other people who worked with him in the context of cybernetics. We could think of John von Neumann, or Warren McCulloch, for example. They were specialized people with good grounds in other sciences, who believed that, as Campbell (2005) states, interdisciplinary training can't have "Leonardesque aspiration".

Interdisciplinary groups

With regards to the concrete research, Wiener's requirements were for very small groups, made up by two-three, maximum four people. He worked at the wartime project on anti-aircraft predictors with an engineer, Julian Bigelow, and two assistants. They worked on electromechanical devices ("predictors", or "directors") to predict the course of enemy aircrafts to help the anti-aircraft gunners. In this context, Wiener played the role of a mathematician, creating the fundamental "[t]heory of the prediction of stationary time series", credited both to him and to the Russian mathematician A. N. Kolmogorov, and introducing statistical methods into control and communications engineering. From 1945 to 1950, that is, at the time of the Macy Conferences, he carried out a project with Arturo Rosenblueth (a neurophysiologist), Walter Pitts (a logician), and Garcia Ramos (a physician). Even in this case, Wiener worked essentially as a mathematician. Ramos told that "Wiener would come to the laboratory and watch them do their experiments, throwing in an idea or two. Then, following further talk on the results, Wiener would retire to work out a mathematical explanation" (Ramos, interviewed by Masani 1990: 206). This operative attitude in the interdisciplinary work group was coherent with the above-mentioned personal interdisciplinary ideal. As Wiener (1948: 9) explains:

The mathematician need not have the skill to conduct a physiological experiment, but he must have the skill to understand one, to criticize one, and to suggest one. The physiologist need not be able to prove a certain mathematical theorem, but he must be able to grasp its physiological significance and to tell the mathematician for what he should look.

Wiener's predilection for very small operative groups was upstream during World War II and the years of the Cold War only if compared for instance with the Los Alamos Project, or

the SAGE Project. It was inspired by two reasons. The first one was the consideration that the "big science" reduced the freedom and the sense of responsibility of the scientists (see e.g., Wiener 1964: 231–232). The second reason was of a more personal nature: a bigger work group would have been unfit to Norbert Wiener's temper and research style. In both aspects, von Neumann, for example, had really different attitudes. He had a much less strict philosophy of responsibility in science, and was a kind of scientist – like Fermi, Oppenheimer, or Vannevar Bush – who knew how to put a large group of people to work.

Another kind of interdisciplinary collaboration, often experienced by Wiener, were seminars composed by, say, 20-25 people, offering the chance of a very intensive interaction. These seminars were very important in his intellectual itinerary, in particular Royce's interdisciplinary seminar on scientific method, which he attended at Harvard between 1911 and 1913, the so-called "dinners" organized in the 1930s by Rosenblueth at the Harvard Medical School. Even the postwar-time Macy Conferences on Cybernetics belong to this type of interdisciplinary collaboration. They were a series of two-day-long meetings which lasted for seven years, from 1946 to 1953. All of these three experiences are strongly emphasized in the introduction in Wiener (1948).

Wiener's confidence in scientific hybridization

Wiener's great confidence in interdisciplinarity was particularly based on his personal experience of the heuristic potential of hybridization among different traditions of thinking. This was already typical of his individual work. Among several possible examples, two are particularly significant. The first is referred to in his very relevant early mathematical work on Brownian motion, at the very beginning of the 1920s. Wiener (1954: 9–10) tells his results in the following way:

> Gibbs had to work with theories of measure and probability which were already at least twenty-five years old and were grossly inadequate to his needs. At the same time, however, Borel and Lebesgue in Paris were devising the theory of integration which was to prove apposite to the Gibbsian ideas. […] Lebesgue […] had neither the sense of physics nor an interest in it. […] I believe that I myself, in 1920, was the first person to apply the Lebesgue integral to a specific physical problem – that of the Brownian motion.

Wiener was firmly convinced that by putting together different, maybe very distant, scientific traditions it was possible to obtain a mutual growth in science. This approach became his most frequent method of doing research. Very suggestively, one can trace the same approach as far back as the first page of his wartime book nicknamed "Yellow peril", in which he presented the Wiener(–Kolmogorov) theory of prediction. As Wiener (1949: 1) stated:

> This book represents an attempt to unite the theory and practice of two fields of work which are of vital importance in the present emergency, and which have a complete natural methodological unity, but which have up to the present drawn their inspiration from two entirely distinct traditions, and which are widely different in their vocabulary and the

training of their personnel. These two fields are those of time series in statistics and of communication engineering.

The sense of "hybridization" we are using here to represent Wiener's approach is stronger than that currently used in anthropology to speak of a "melting pot", which inevitably entails an ever imperfect process of fusion. The outcome of the fusion envisaged by Wiener (1949) was a coherent theory, a generalized theory.

The origins of cybernetics during World War II

The cybernetics program sprang directly along this course. A first complete picture of the tumultuous field that Wiener decided to call cybernetics, choosing the title for Wiener (1948), appeared in a meeting of January 1945 at Princeton. Among the participants, there were von Neumann, Wiener, McCulloch, Pitts, and Lorente de Nó. Following Heims (1991), we shall refer to these people as to the "group of Princeton".

In Wiener's opinion, which was substantially shared by other members of the group of Princeton, the paradigm of cybernetics was epitomized by two threads of work. On the one hand, the research of Wiener and Bigelow on predictors, with its prosecution in the neurophysiological field, as shown in the article by Rosenblueth, Wiener, and Bigelow (1943), where the purposive behavior in living organisms is studied in parallel with machines by means of the theory of automatic control. On the other hand, the research on digital computers, which Wiener had began in 1940 sending to the US National Defense Research Committee a *Memorandum* on digital computers to solve partial differential equations. He continued this research in the context of a more or less informal collaboration with von Neumann, as shown by Montagnini (2005). This second thread had a physiological side as well, in particular the article by McCulloch and Pitts (1943) on neuronal nets, finalized to explain the behavior of the human brain, which was used by von Neumann in the design process of the first general purpose computer, the Edvac.

In both of these cases, the role of the interdisciplinary collaboration was clear and effective: the experts in "intelligent" machines (digital computers and analogical controlled devices) were able to help the neurophysiologists, and vice versa. Ideas, results and future perspectives emerging from that work were discussed during a secret meeting at Princeton, 6–7 January 1945, organized by Wiener and von Neumann. The program – written by Wiener – contained the proposal for the creation of a small scientific society to discuss an emerging generalized theory of communication embracing, apart from the classic subjects of communication engineering (radio, telephone, radar, etc.), as well as computing machines, control devices, and the "communication and control aspects of the nervous system" (Wiener's letter to Goldstine, December 22, 1944, as cited in Goldstine 1980: 275). Here is the new science that Wiener (1948) named *cybernetics, or control and communication in the animal and the machine.*

Cybernetics was carried out in very circumstantial researches, based on detailed experiments and rigorous mathematical work. From 1945 to 1950, the group of Wiener and Rosenblueth worked on cardiac flutter, clonus and nerve conduction. We could find analogous style

in McCulloch's research work. Von Neumann's main concrete work in cybernetics consisted in the design of "general purpose computers".

Difficulties entailed by interdisciplinarity were probably present in these narrower contexts, but in a very controlled form. On the contrary, they emerged in so far as specific researches needed to be put together in forming a wider theoretical picture. And they became even more relevant when the Princeton group met socio-human scientists at Macy conferences.

The new language of cybernetics

Enthusiastically, Wiener exclaims in his memoirs that, at the Princeton meeting, cybernetics was born, together with a new language. As he writes (Wiener 1964: 269; italics added):

> [all the participants] were interested in the storage of information to be used later, and all of them found that the word *memory* (as used by the neurophysiologist and the psychologist) was a convenient term to cover the whole scope of these different fields. All of them found that the term *feedback,* which had come from the electronics engineer and was extending itself to the servomechanism man, was an appropriate way of describing phenomena in the living organism as well as in the machine. All of them found that it was convenient to measure information in terms of numbers of yeses or noes, and sooner or later they decided to term this unit of information the *bit.* This meeting I may consider the birthplace of the new science of cybernetics, or the theory of communication and control in the machine and in the living organism."

Observing in passing that this is still "our" language, one can legitimately wonder about the kind of language it was. Actually, scholars studying interdisciplinarity, e.g., Fuller (1993), and Klein (1996, 2005) have stressed that it is a common event that, during interdisciplinary interaction, an interlingua arises. Thagard (2005) has paralleled Paul Smolensky idea about the existence of a sort of "pidgin" among the present cognitive scientists with the analogous Galison's (1997) hypothesis about the presence of "trading zones" in science, a sort of no man's lands where different subcultures interact, for instance intercultural situations studied by anthropologists. Nevertheless, for Smolensky the present interlingua in cognitive science is not a true language: it appears to be more similar to a "pidgin" than to a "creole". Galison (1997) holds the same views about science in general. In his opinion, based on historical studies on microphysics researches during World War II, an interlingua sprung then, enabling scientists to interact exquisitely about "local detail" but "without global agreement".

Although cyberneticists aspired to a true new language, it is certain that at the Princeton meeting they only used a pidgin. Actually, still at present, the notions of "feedback" (see Richardson 1991), and "information" (see for instance de Luca 2006, and Termini 2006a, 2006b) are polysemic and controversial, even in the context of the most mathematized sciences. Problems increase with the enlargement to neurosciences, not to mention the socio-human sciences.

The "Oregon epistemology"

The program of early cybernetics suffered from naivety in neglecting a number of "invisible" differences underlying various traditions, while dealing with apparently similar notions. I believe that in order to understand some of these problems, it is useful to reflect on the Oregon metaphor, suggested by Wiener in *Cybernetics:*

> Specialized fields [in science] are continually growing and invading new territory. The result is like what occurred […] [in] Oregon […] – an inextricable tangle of exploration, nomenclature, and laws. There are fields of scientific work […]; in which every single notion receives a separate name from each group, and in which important work has been triplicated or quadruplicated, while still other important work is delayed by the unavailability in one field of results that may have already become classical in the next field. (Wiener 1948: 8)

As stressed by Lepschy (1998), we have to consider closely the situation of Oregon at the beginning of the 19th century, that was quite unlike the present. Today, more neutrally, historians usually call that region the "Pacific Northwest". It was a wide land bound in the west by the Pacific Ocean, in the north by the still Russian Alaska, and in the south by Mexico. It had been colonized by the French, German, Spanish, English, Russian, not to mention the Amerindians. For a long time, the United States called it Oregon, and the United Kingdom – who claimed the northern part of it – used to call it British Columbia.

Wiener and Rosenblueth thought that the situation in their contemporary science was similar, and therefore it was like a mine, just because of the opportunities of hybridizations. But encounters among different traditions also involve various kinds of troubles. There is certainly a bias in considering the equation "map = territory". But let me contradict the famous *dictum* of Alfred Korzybski. One ought to consider that in a certain sense the contrary is true as well: "The map 'is' the territory". To get to America, Columbus used a map that measured in a wrong way the distance from the Far East and Portugal, and another one by Marco Polo's *Milione.* As a result, he took the Antilles for Japan, and Arawak for India. Symmetrically, Arawak, following some legends, took the newcomers with their arquebuses for a sort of gods. Different maps are like different glasses, therefore, in Oregon the same towns could have different laws, and, in addition, some people could see things that other people saw in a different way, or maybe could not even see at all. This is very true not only for intercultural situations but also for interdisciplinary collaboration. In this sense, I agree, except for his pessimistic conclusions, with Bauer (1990) in stressing the specific *forma mentis,* the procedures, etc. implicit in different disciplines.

Interdisciplinary difficulties at the Macy Conferences

At the Macy conferences on Cybernetics, the Princeton group met another group of scientists, made up by neurophysiologists, psychiatrists, psychologists, an economist, a sociologist and two anthropologists, all more or less hinging around the Macy Foundation (see Heims 1991, and Dupuy 1999).

The first problem they met concerned the same theme of the meetings. The conferences had started from the wish of social and human scientists to know better the concept of circular causality and feedback, and how these notions were treated by Wiener and his collaborators, in order to verify their applicability to human psychology and society. On the contrary, the Princeton group was essentially interested in continuing the path of the Princeton meeting: to study in parallel organisms and "communication machines", whilst the feedback was for them nothing more than a conceptual tool among others. The difference is shown by the subtitles of the meeting transactions: *Cybernetics: Circular causal and feedback mechanisms in biological and social systems,* when compared with that of Wiener (1948).

The "pidgin" of the Princeton meeting was further overburdened by different "influential metaphysics" that worked in the background, giving different importance to similar notions. As shown by Montagnini (2008), the group hinging around the Macy Foundation was pervaded with holistic and relational ideas, while the other tended to be more mechanistic and reductionist. Difficulties increased by the lack of mathematical culture among "soft" and "hard" scientists. Although the "mathematical divide" in fact became a sort of scapegoat due to its visibility. In the meantime, the entire world was slipping into a global conflictual vicious circle: the Cold War. And this warmed up the human atmosphere of the conferences, putting on the table new and huge ethical problems.

Convergences, but...

The Wiener's (1948) publication took place between the 5th and the 6th meetings, marking a partition between the first five conferences and the following five ones. In fact, the Princeton group's program got the best of the more vague soft scientists' expectations. Psychologists were co-opted in the animal–machines study under the aegis of communication, focusing more and more on brain. During the Macy conferences, a big effort toward the negotiation of meaning was constructively done. Some ambiguities were clarified. Some people gave up, while scientists like Bateson or Mead accepted the idea of finding a compromise between their holistic views and their scientific needs in order to understand human behavior (see Montagnini 2007). Bateson still accepted the idea to use machines as models for mental phenomena; see for instance the way in which Bateson (1987) achieves the double bind theory of schizophrenia (125–126). On the other hand, members of the Princeton group assimilated some of the holistic spirit, as shown for instance by the two chapters added in Wiener (1948, 1961), where the concept of self-organization emerges.

During the 1950s and 1960s, cybernetics with its language, concepts and metaphors became popular, still modeling the mental atmosphere of the Cold War. Nevertheless, a more in-dept theoretical reflection came to a halt. Was interdisciplinarity the stumbling block of this event? Our analysis does not justify this conclusion. Bigelow noted that the same Princeton group broke up due to "a clash of personalities among organizers" (Aspray 1990: 302). This conflict pertained to moral problems, in connection with the atomic bombing of Japan, the Cold War tensions, and also with more complex psychological reasons. The fact is that Wiener, von Neumann, and McCulloch did not cooperate anymore. Afterwards, the different influential metaphysics underlying the Macy Conferences seemed to become like many "nu-

clei of aggregation" for new approaches. How and why this happened deserves to be studied in depth.

References

Aspray, W. (1990) The origins of John von Neumann's theory of automata. In: J. Glimm, J. Impagliazzo, and I. Singer (eds.) *The Legacy of John von Neumann.* Providence: AMS, 289–309.

Bateson, G., and Bateson, M. C. (1987) *Angels Fear: Towards an Epistemology of the Sacred.* New York: Macmillan.

Bauer, H. H. (1990) Barriers against interdisciplinarity. *Science, Technology, and Human Values* 15 (1), 105–119.

Campbell, D. T. (2005) Etnocentrism of discipline and the fish-scale model of omniscience. In: S. J. Derry, C. D. Schunn, and M. A. Gernsbacher (eds.), *Interdisciplinary Collaboration: An Emerging Cognitive Science,* 3–21.

Derry, S. J., Schunn, C. D., and Gernsbacher, M. A. (eds.) (2005) *Interdisciplinary Collaboration: An Emerging Cognitive Science.* Mahwah: Lawrence Erlbaum.

Dupuy, J. P. (1999) *Aux origines des sciences cognitives.* 2nd ed. Paris: La Découverte.

Fuller, S. (1993) *Philosophy, Rhetoric, and the End of Knowledge.* Madison: University of Wisconsin Press.

Galison, P. (1997) *Image & Logic: A Material Culture of Microphysics.* Chicago: University of Chicago Press.

Goldstine, H. H. (1980) *The Computer from Pascal to von Neumann.* Princeton: Princeton University Press.

Heims, S. J. (1991) *The Cybernetics Group.* Cambridge (MA): MIT Press.

Klein, J. T. (1996) Crossing Boundaries: Knowledge, Disciplinarities, and Interdisciplinarities. Charlottesville: University Press of Virginia.

Klein, J. T. (2005) Interdisciplinary teamwork: The dynamics of collaboration and integration. In: Derry, S. J., Schunn, C. D., and Gernsbacher, M. A. (eds.) *Interdisciplinary Collaboration: An Emerging Cognitive Science,* 23–50.

Lepschy, A. (1998) Interdisciplinarità e metadisciplinarità dai punti di vista dell'ingegneria dell'informazione e della cibernetica. *Accademia e Interdisciplinarità I: Saggi.* Padova: Accademia Galileiana di Scienze, Lettere ed Arti.

de Luca, A. (2006) Some reflections on cybernetics and its scientific heritage. *Scientiae Mathematicae Japonicae* 64 (2), 243–253.

Masani, P. R. (1990) *Norbert Wiener: 1894–1964.* Basel: Birkhäuser Verlag.

McCulloch, W. S., and Pitts, W. (1943) A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics* 5: 115–137.

Montagnini, L. (2005) *Le armonie del disordine. Norbert Wiener matematico-filosofo del Novecento.* Venezia: Istituto Veneto di Scienze, Lettere ed Arti.

Montagnini, L. (2007) Looking for "scientific" social sciences: The Macy Conferences on Cybernetics in Bateson's itinerary. *Kybernetes* 36 (7–8), 1012–1021.

Montagnini, L. (2008) Philosophical Approaches towards Sciences of Life in Early Cybernetics. In: Ricciardi L. M., Buonocore A., and Pirozzi, E. (eds.) *Collective Dynamics on Competition and Cooperation in Biosciences. Proceedings of BIOCOMP2007. Vietri Sul Mare 24–28/9/2007.* New York: AIP, 11–17.

Richardson, G. P. (1991) *Feedback Thought in Social Science and Systems Theory.* Philadelphia: University of Pennsylvania Press.

Rosenblueth, A., Wiener, N., and Bigelow, J. (1943) Behavior, purpose and teleology. *Philosophy of Science* 10: 18–24.

Termini, S. (2006a) Remarks on the development of cybernetics. *Scientiae Mathematicae Japonicae* 64 (2), 461–468.

Termini, S. (2006b) Imagination and Rigor: Their interaction along the way to measuring fuzziness and doing other strange things. In: Termini, S. (ed.) *Imagination and Rigor.* Milano: Springer, 157–173.

Thagard, P. (2005) Being interdisciplinary: Trading zones in cognitive science. In: Derry, S. J., Schunn, C. D., and Gernsbacher, M. A. (eds.) *Interdisciplinary Collaboration: An Emerging Cognitive Science,* 317–339.

Wiener, N. (1948, 1961) *Cybernetics, Or Control and Communication in the Animal and the Machine.* 1st ed., 2nd ed. with two chapters added. Cambridge (MA): MIT Press.

Wiener, N. (1949) *Extrapolation, Interpolation, and Smoothing of Stationary Time Series, with Engineering Applications.* New York etc.: Wiley & Sons. First published in 1942 as a classified report to Section D2, NDRC.

Wiener, N. (1950, 1954) *The Human Use of Human Beings.* London: Eyre and Spottinswoode (revised ed.). Garden City: Doubleday.

Wiener, N. (1964) *I am a Mathematician.* Cambridge (MA): MIT Press.

DISPOSITIONS OR MECHANISMS: ON THE QUESTION OF THE SUBJECT MATTER OF THE PHILOSOPHY OF PSYCHOLOGY

Gabriele M. Mras

In the face of developments in the philosophy of psychology, one might be tempted to agree with a judgment as harsh as Wittgenstein's:

> [...] when we philosophize [...] we behave like savages, primitive people, who hear the expressions of civilized men but put a false interpretation on them, and then draw the queerest conclusion from it. (Wittgenstein 2009: §194)

One does not have to be a Wittgensteinian to endorse that verdict. If the following checklist is correct, the various "-isms" offered in the course of that development suffer from a deficiency that is remarkable in two ways. The reason originally given for opposing behaviorism was that behaviorism "leaves out the mind". However, if it is true that "functionalism leaves out the mind", type-physicalism and token-physicalism "leave out the mind", and "cognitivism leaves out intentionality" (Searle 1992), then every particular account ends up making the very same mistake. What has been left out is not just something without which "the psychological" would not be complete, but something essential for there being a subject matter of psychology at all. This situation calls for not another new account in the same tradition, but for a diagnosis, or explanation.

I think that analyzing the historical development of the philosophy of psychology can be of help here. Given that the views in question – behaviorism, functionalism, cognitivism, etc. – were developed in order to overcome what in the end turned out to be true of them, too, the question "What is responsible for this fact?" gives rise to the following riddle: How is it possible that these views all share the very characteristic they object to in each other? Take for example cognitivism: How is it explainable that cognitivism shares with behaviorism what is commonly labeled "reductionism", if reductionism is precisely what cognitivism criticizes behaviorism for? To answer this question I want look at the attitude of both behaviorism and cognitivism towards the following question: "What justification, if any, can be given for the claim that one can tell, on the basis of a person's behavior, that he is in a certain mental state?" (Chihara 1973: 137).

The conceptual status of behaviorism

Let us go back to the very theory which all subsequent theories opposed: "In the beginning was behaviorism" (Searle 1992: 33). Behaviorism, the "scientific metatheory that dominated

psychology between 1913 and 1960" (Baars 1986: 5) proposed an understanding of psychology determined by the requirement that only what is *objective* and scientifically confirmable can rightly be seen as the subject matter of any psychology that deserves to be called "scientific". For this reason, "behavior" became the essential concept of this approach. Publications in psychology had in those days already in their title "behavior", or any other terms related to the behaviorists' understanding of it: Watson's *Psychology as the Behaviorist Views it,* published in 1923, Guthrie's *Psychology of Learning* of 1935, B. F. Skinner's *The Behavior of Organisms* published in 1938, and *Schedules of Reinforcement* by Skinner and C. B. Ferster, published in 1957, the same year in which *Verbal Behavior* was published. The proponents of behaviorism made it completely clear that by what they meant by "objective" and "scientific" was not to pursue one's own investigation by exhibiting the virtues of "detachment and clear thinking". The requirement to accept nothing that can be proven scientifically was also constitutive for what *behavior* was taken to be.

> Behaviorism [...] is an objective approach to psychology. By this is meant not merely that it is characterized by detachment and clear thinking, of the kind we so much admire in the scientist, who, skeptical seeker after truth, believes only what he can prove. There is in behaviorism an implied opposition to introspectionism. (Stephenson 1953: 110)

> The behavior of an organism is not [...] an object [...] for introspection. It is a process, a continuous change. [...] A science must achieve more than a description of behavior as an accomplished fact. It must predict future courses of action; it must be able to say that an organism will engage in behavior of a given sort at a given time. (Skinner 1961: 70)

The main complaint about the behaviorists' method was, and still is, that it has as its consequence that wide parts of psychology would have to be "dropped without hope of a substitute" (Fodor 1965). But philosophers of psychology in the 1960s understood by such "parts" not what Titchener's introspectionism would have yielded as objects of a psychological theory – consciousness' elements or properties (Titchener 1897). The philosophers who reacted against behaviorism's assumptions missed in the behaviorist explanation of behavior something very ordinary: the use of everyday psychological concepts like "belief" and "desire" This gave rise to the aim to "implement" these concepts – contrary to behaviorism's neglect (of them) – as conditions of formulating psychological generalizations or laws. Thus, after "behaviorism was given a hearing for fifty years" (Koch 1999: 56), philosophers, psychologists, and computer scientists began to develop views that were in retrospect labeled as "cognitive revolution". The objections raised against the various kinds of behaviorism were that by (1) ignoring beliefs and desires (2) causal explanations were made impossible which led to (3) a restriction of the subject field of psychology. The arguments offered for a *refutation* of behaviorism were basically two-fold: (a) because of (1) and (2) behaviorism restricts psychology in a way that makes it even impossible for behaviorism itself to pursue its purpose, i.e., to predict which "future action" "an organism will engage in" "at a given time (Hamlyn 1953; Putnam 1975; Fodor 1965), and that as another consequence of (1) and (2b) behaviorists' explanations of behavior were helplessly circular (Chisholm 1957; Geach 1957; Fodor 1965). Argument (a) notoriously invokes the idea of a criterion of identity of behavior and even movements:

> [...] but the behavior which we call "posting a letter" [...] involves a very complex series of movements. [...] No fixed criteria can be laid down which will enable us to decide what series of movements shall constitute "posting a letter." (Hamlyn 1953: 134)

> The requirement that psychology has to deal with movements alone is proved to be a mistake by the impossibility of construing a criterion of identity of movements. (Fodor 1965)

This criticism clearly targets the empiricism of behaviorism and raises doubts whether the object of behaviorism's investigation is clearly defined. What behaviorism understands by observable behavior is the question. However, the invoked "identity criterion of movements" is not apt for raising the issue of behaviorism's *empiricism.* For sure, if it were the case that a description of behavior would require everything mental to be deleted, nothing would remain as the possible object of psychological investigations, then occurrences, or "movements"; in consequence, no criterion for distinguishing kinds of behavior – e.g., "posting a letter", or "kicking a ball" – could be given. However, the introduction of the idea of such a criterion does not have the effect Fodor thinks it has, because whether something follows from the requirements of such a criterion depends on whether behaviorism's restriction of the subject matter of psychology has the consequence as assumed. And here, more needs to be said than that one cannot both regard psychology as a distinct enterprise and refuse to distinguish it from all other sciences whose subject matters could be described as behavior. It is worth noting, too, that neither Watson, nor Skinner, nor any other behaviorist proposed such an impoverished conception of "behavior", and none of them understood the exclusion of beliefs and desires as explanatory relevant factors as to express a denial of the *existence* of mental states. To bring out the impossibility of behaviorism, it must be shown *how* its conception of what observation provides as "behavior" ultimately ends in the assessment that "observables" are "happenings", "occurrences", or "movements" alone.

The mentalists'/cognitivists' criticism so far suggests that what is defended by behaviorism as a necessary condition of any scientific psychology counts for cognitivism as denying its very possibility. For behaviorism, the philosopher of psychology is on the wrong track if he tries to explain behavior by an appeal to mental events:

> Under the influence of a contrary philosophy of explanation, which insists upon the reductive priority of the inner event many brilliant men who began with an interest in behavior [...] have turned instead to the study of physiology. (Skinner 1950: 325–326)

Cognitivism, however, regards this way into the realm of "the inner" as necessary for understanding behavior as an expression of what human beings believe and want. The reason for these different assessments lies in the different understandings by cognitivism and behaviorism of what is available in *observable behavior.*

For a behaviorist, of the kind of Watson and Skinner, the answer to "What is observable behavior?" depends on the method that helps to define behavior. Thus, "observable behavior" and "being the consequence of external conditions" are interchangeable in meaning. It is primarily this view of behavior, i.e., the concern to formulate contingencies whose variables range over *external* factors alone, which is responsible for behaviorism's deviation from

"folk psychological" explanations. Nevertheless, it is true that in the course of behaviorism's defense of this "instrumentalist" view, the applicability of "causal relations" is discussed in a way that shows itself obliged to the verificationism by logical empiricists. Whether it is verificationism that allows behaviorists to exclude psychological states in explanations of behavior, or a certain understanding of causal explanations which had this result is ultimately not decidable. Skinner's reasoning is a good example of how causal relations were understood:

> When we attribute behaviour to a [...] mental event, real or conceptual, we are likely to forget that we still have the task of accounting for the neural or mental event. [...] what began as the task of accounting for learned behaviour becomes the task of accounting for expectancy. The problem is at least equally complex and probably more difficult. (Skinner 150: 32)

According to this reasoning, the explanation of behavior A by the mental event B only shifts the question from "How did A come about?" to "How did B come about?". Since the mental event B has a different content from behavior A, this leads to the question "How did A arise from B?" However, this question cannot be answered because all that is available in order to say something about B is just event A. The conclusion that mental events as causes of behavioral events are just postulates is close, and based on the assumption that behavior A is the only indicator of the occurrence of the mental event B but necessarily different in content. Emphasizing the quality any explanation "from" mental events would have, Ryle concludes:

> This was indeed the mistake of the old Faculty theories which construed dispositional words as denoting occult agencies or causes, i.e. things existing, or processes taking place, in a sort of limbo world. (Ryle 1949: 119f.)

To sum it up: The behaviorist objection to mentalism is that it simply invents mental entities as allegedly causally relevant factors in an explanation of behavior. Its explanations are therefore fictions. This conclusion is reached because for behaviorism the question "How did A arise from B?" is a question about how content can have a causal efficiency by itself. Because this question cannot be answered, but both questions are put on the same level, *behavior* is not to be seen as an "expression" of a person's having particular beliefs and desires. If this (mistaken) view of what would be required to show beliefs to be causally relevant is supported by an empiricist's argument of what is observable, then what is reached is now the reductionist view that Fodor ascribes to behaviorism. But then it would also be senseless to pursue the goal of *producing* dependencies.

Therefore, if one still wants to find it intelligible to *influence* behavior, one has to understand "behavior" in two ways: on the one hand, as a consequence which is described as a relation to its (external) conditions, and on the other hand, as what *enables* the (external) conditions to have that consequence. The term "disposition" is a term that is used to express both of these aspects.

Therefore, when Fodor asked the behaviorista about their criterion of the identity of movements, he gave away what actually shows the inconsistency of behaviorism. Instead of using this question in order to demonstrate that behaviorism always *presupposes* what it wants to have eliminated, Fodor aims at answering this question of the conditions of the

right criterion himself. This is so because he thinks that formulating a criterion of the *content* of observations of behavior represents an indirect proof of the truth of cognitivism; and this indirect proof will involve that the question "What justification, if any, can be given for the claim that one can tell, on the basis of a person's behavior, that he is in a certain mental state?" starts from a "level" that cognitivism criticized behaviorism for operating on.

The status of mentalistic concepts

When one reads programmatic statements of the kind of mentalism that emerged in the 1960s, it looks as if all of behaviorism's postulates were rejected:

> It seems perfectly obvious that what's needed to construe cognitive processes is precisely what behaviorists proposed to do without; causal sequences of mental episodes and a 'mental mechanics' to articulate the generalizations that such sequences instantiate (Fodor 1981: 6)

For behaviorists as well as for some of the philosophers in the tradition of analytic philosophy, asking "How does the mind work?" or "What is in the mind's machinery?" indeed commits one to making the "mistakes of the old faculty theory". Gilbert Ryle wrote decades after the publication of *The Concept of Mind* to Dennett:

> 'Cognitive psychology' sounds to me like the later days of phlogiston! It looks as if Fodor took unexamined some bogus notion of 'internal' and then excogitate hypotheses about the ways in which postulated things, happenings, etc. in this 'internal' region can go proxy for things… (Ryle 1976)

But given statements like "[...] what the behaviorists seemed to have got right – contra the identity theory – was the relational character of mental properties" (Fodor 1981: 9), one cannot but recognize the striking similarities between mentalism and behaviorism. This is not just to doubt whether the label "cognitive *revolution*" is really used adequately here given that the various research programs united under the name "cognitive sciences" look less like an expression of a paradigm shift but could be understood too as a return – away from instrumentalist views – to conceptions of psychological explanation pursued before the "heydays" of behaviorism (see Mandler 2002; Greenwood 1999; Cohen-Cole 2005; Boden 2006). The purpose that is to be served by the developing of "mechanical models of organisms" – "Isn't this, in a sense just what psychology is about?" (Putnam 1975: 435) –, and the basis of this "hypothesis" itself is derived from behaviorism: namely, the picture of the psychological as something "inside" the subject and the behavior as something "outside of it". So it is only because, in contrast to the assumed method of behaviorism, cognitivism seeks to verify what can be said about "outside" by its relation to "inside", that is the mental. What this kind of "verification" then requires is to show that through the workings of the mind, uninterpreted data – mere movements of bodies – are "transformed" into behavior whose content is more than these occurrences. The conviction that the behaviorist's aim to

formulate "laws of behavior" as better pursued in this framework is recognizable. In the *kind* of circularity charge, Fodor raises against dispositional explanations, too:

> Suppose 'John took aspirin because he had a headache' is true if conjunction C holds: C: John was disposed to produce headache behaviors and being disposed to produce headache behaviors involves satisfying the hypothetical if there are aspirin around, one takes some, and there were aspirin around. So, C gives us a construal of 'John took aspirin because he had a headache.' But consider that we are also in want of a construal of statements like 'John was disposed to produce headache behaviors because he had a headache.' [...] in these cases it seems unlikely that the putative mental causes can be traded for dispositions; [...] We cannot, for example, translate 'John had a headache' as 'John had a disposition to produce headache behaviors. (Fodor: 1981: 4f)

What is obvious in this discussion of "a construal of statements like 'John was disposed to produce headache behaviors because he had a headache" is that behaviorism is treated as if it were, it had to be, a version of representationalism. This is why dispositional explanations are used in a way that partly allows to "translate" statements about mental states into statements about behavioral dispositions (in the first sentence), but not "so far" that the disposition itself could be seen as what different kinds of pain behavior have in common: Because if we were to substitute "to feel pain" by "a disposition to feel pain" in a sentence, where pain is not regarded as the cause of headache behaviors, but as the disposition to pain behaviors, then this would "obviously" result in, no, not just a tautology, but in a sentence that states something *different.* But with the same right, one could use this procedure in reverse: since (1) the sentence "John was disposed to produce headache behaviors because he had a headache" needs to be reconstructed, and (2) "disposed to produce headache behaviors" is supposed to be equivalent to "having a headache", the sentence "John had a headache because he had a headache" is also of considerable interest. The point is that neither (1) nor (2) have anything to do with the position of behaviorism, and that hence Fodor's entire reconstruction presupposes the task to show what cognitivism cannot reformulate in terms of a behaviorist theory. For behaviorism, "feeling pain" and "the disposition to pain behavior" are not thought to be *equivalent* because what is meant by the expression "to feel pain" is not a scientifically verifiable fact; and when associations are made between observable pain behavior and the expression "to feel pain", this entails an *assumed* (weaker than inferred) state *from* pain behavior.

What Fodor uses here against dispositional explanation is that behaviorism is basically confronted with the same problem which behaviorism attributes to mentalism: either by "disposition" nothing is added to what is "observable": that behavior is said to "follow" from a "disposition" supposedly states no more than that there is a relationship between consequences and their (external) condition. If, though, "having a disposition" expresses the idea (2) of a sine qua non (condition) of the possibility of behavior (= an external condition having a consequence), a modified version of a dualistic explanation of actions from a separated "place" is reborn. This is why Fodor, Putnam, and the others can claim that their mechanical models are descriptions of dispositions, required by the explanatory deficiency of explanations by dispositions.

One only has to appreciate the characteristics of the mental as *not* being entailed in behavioral expressions in order to acknowledge that this way to show behaviorists' explanations to be

circular would run counter the cognitivists', mentalists', and representationalists' aim. If beliefs and desires are characterized solely as being separated from the interpreter, insofar as autonomous and unobservable in relation to the interpreter, then any "externalization" would blur the character of beliefs and desires as being *mental.* The non-dispositional characterization of beliefs, desires, as "unobservable causes" would make it impossible to explain what one did by appealing to what one thought and wanted. It is now the whole point of Fodor's remarks that it is a mistake. One interprets it both as criticism of this observational basis as well as its affirmation: an "observational basis" that disputes "everyday" observations must be adopted in order to show it to be justified:

> [...] it is sometimes claimed that, at least in some cases, no inference from behavior to mentals is at issue in psychological ascriptions. Thus, we sometimes see that someone is in pain, and in these cases we cannot be properly said to infer that he is in pain. However, the sceptic might maintain against that argument that it begs the question. [...] the sceptic can argue that what is required in the case of another's pain is some justification for the claim that, by observing a person's behavior, one can see that he is in pain. [...] To hold that the sceptical premise is false is ipso facto to commit oneself to some version of logical behaviorism. (Fodor 1981: 36)

But this way of justifying "the claim that one can tell, on the basis of a person's behavior, that he is in a certain mental state" (Chihara 1973: 137) brings mentalism into conflict with everyday explanations of behavior. For desires and beliefs are viewed here as "theoretical constructions" whose truth is to be demonstrated and must not be assumed. The skepticism that is adopted for methodological reasons, however, requires to entertain two attitudes towards the question how one "can tell, on the basis of a person's behavior, that he is in a certain mental state", which oppose each other: on the one hand, a mental state is something that is in its very nature distinct from observable behavior; on the other hand, it is exactly through what is understood as being different in its "essence" that the existence of mental states is thought to be provable.

For behaviorism, the division of "the psychological" in an "inner" and "outer" part was a reason to declare the investigations of the mutual dependencies of the "outer" aspects to represent an autonomous subject matter. Philosophers like Putnam and Fodor aimed at criticizing behaviorism in a way that the critique itself would open up a view to a new science of the mind – "cognitive science". This science was meant to investigate the relationship of the mind to the physical world *both* from a point of view of outside and from a point of view of inside – for the purpose to show that conditions of interpreting something as the "behavior" of human beings are satisfied only by the assumption that this "behavior" is a manifestation of thinking subjects – what required to adopt a view that in turn resulted in what originally was to be avoided: the widely-criticized reductionism in the philosophy of psychology.

References

Baars, J. B. (1986) *The Cognitive Revolution in Psychology.* New York: Guilford Press.
Boden, M. (2006) *Mind as Machine: A History of Cognitive Science.* New York: Oxford University Press.

Chihara, Ch. (1973) Operationalism and ordinary language revisited. *Philosophical Studies* 24 (3), 137–157.
Chihara, Ch., and Fodor, J. A. (1965) Operationalism and ordinary language. *American Philosophical Quarterly* 2 (4), 281–295. Reprinted in: J. A. Fodor (1981) *Representations: Philosophical Essays on the Foundations of Cognitive Science.* Cambridge (MA): University of Cambridge Press.
Chisholm, R. (1957) *Perceiving: A Philosophical Study.* Ithaca: Cornell University Press.
Cohen-Cole, J. (2005) The reflexivity of cognitive science: The scientist as model of human nature. *History of the Human Sciences* 18 (4), 107–139.
Fodor, J. A. (1965) Explanations in psychology. In: M. Black (ed.) *Philosophy in America.* Ithaca: Cornell University Press, 161–179.
Fodor, J. A. (1981) Representations: Philosophical Essays on the Foundations of Cognitive Science. Cambridge (MA): University of Cambridge Press.
Geach, P. (1957) Mental Acts: Their Content and Their Objects. London: Routledge & Kegan.
Greenwood, J. D. (1999) Understanding the 'cognitive revolution' in Psychology. *Journal of the History of the Behavioral Sciences* 35 (1), 1–22.
Hamlyn, D. W. (1953) Behaviour. *Philosophy* 28 (April), 132–145.
Koch, S. (1964). Psychology and emerging conceptions of knowledge as unitary (1–39). In: T. W. Wann (ed.) *Behaviorism and Phenomenology.* New York: University of Chicago Press. Page references to the republication are in: S. Koch (1999) *Psychology in Human Context. Essays in Dissidence and Reconstruction.* Chicago: Chicago University Press, 51–90.
Mandler, G. (2002) Origins of the cognitive (r)evolution. *Journal of the History of the Behavioral Sciences* 38 (4), 339–353.
Putnam, H. (1967) The nature of mental states. In: H. Putnam (1975), *Mind, Language and Reality. Philosophical Papers.* Vol. II. Cambridge (MA): Cambridge University Press, 429–440.
Putnam, H. (1975) *Mind, Language and Reality. Philosophical Papers.* Vol. II. Cambridge (MA): Cambridge University Press.
Ryle, G. (1949) *The Concept of Mind.* London: Hutchinson. London: Penguin Books. Page references are to the 2000 publication.
Ryle, G. (1976) Ryle's last letter to Daniel Dennett. Reprinted in 2002 in: *The Electronic Journal of Analytic Philosophy* 7.
Searle, J. R. (1992) *The Rediscovery of the Mind.* Cambridge (MA): University of Cambridge Press.
Skinner, B. F. (1950) Are theories of learning necessary? *Psychological Review* 57.
Skinner, B. F. (1961) The analysis of behavior. In: B. F. Skinner: *Cumulative Record. A Selection of Papers.* New York: Appleton-Century-Crofts, 70–76. Page references are to the 1972 edition.
Stephenson, W. (1953) Postulates of behaviorism. *Philosophy of Science* 20 (2), 110–120.
Titchener, E. (1897) *An Outline of Psychology.* New York: Macmillan.
Wittgenstein, L. (2009) *Philosophical Investigations.* Revised, 4th edition by P. M. S. Hacker and Joachim Schulte. Oxford: Wiley-Blackwell.

COGNITIVE SCIENCE AS A RESPONSE TO THE CRISIS OF DISCIPLINES

THE HISTORY OF COGNITIVE SCIENCE BETWEEN BIOLOGY AND THE SOCIAL SCIENCES

Andreas Reichelt & Nicole Rossmanith

Introduction

Why take time off doing cognitive science to read or write about its history? Or, for that matter, why take time off researching the history of science to look at current developments in the field, or to talk to the practitioners turned amateur historians, and listen to their partisan accounts of the history of their discipline?

In general, many scientists and historians alike do not seem to bother too much about each other's lines of work. However, in the case of a relatively young field such as cognitive science, the relationship between accounts of its history and current developments is remarkably close, with most, if not all, of the influential accounts produced by people who are active in cognitive science in one way or another. As cognitive science has been an interdisciplinary project from its inception, and has become more and more diverse still since the 1960s, a shared historical narrative is especially significant to the identity of our community and its practitioners.

Therefore, we begin by providing a brief account of the "standard" rendition of the history of the discipline. Next, we survey the developments in the field, and ask when cognitive scientists drew on what kind of historical sources to add to or change the official story. We conclude the first part with our own account of the history of cognitive science, drawn from an outside perspective that views the history of the field in terms of its relationships to other scientific traditions. By focusing on two key neighboring fields of study, the biological sciences on the one hand, and the socio-cultural sciences on the other, it is possible to make sense of many developments in cognitive science, including the crucial new trends of embodied action and situated cognition.

In the second part, we focus on the practitioner historians of cognitive science – from both inside and outside the mainstream – their agendas, and the different use they make of the historical record. Can we not get rid of all this haggling and tinkering with history by turning to proper, objective, disinterested historiography? We present some reasons why this will, alas, not be possible, and give examples of how following current developments in and studying the history of cognitive science can mutually benefit each other.

Part one: Histories of cognitive science – towards an outside perspective

> *Psychology is committed to investigating processes like cognition, perception, and motivation, as historically invariant phenomena of nature, not as historically determined social phenomena. Accordingly, [...] historical studies have as little relevance for current work in the discipline of Psychology as the history of physics has for current work in that science. In both cases, the low status of history depends on an implicit belief in scientific progress. If the historical course of science represents the cumulative improvement of knowledge, then the past simply consists of that which has been superseded. The main reason for bothering with it at all is to celebrate progress, to congratulate ourselves for having arrived at the truth which the cleverest of our ancestors could only guess at. (Danziger 1997: 9)*

Tracing the inside history of cognitive science

There are now several narratives (Bechtel, Graham, and Abrahamsen 1998; Boden 2006; see Varela, Thompson, and Rosch 1991 for a different perspective) supplementing and adding to the standard account given by Howard Gardner (Gardner 1985). Together we refer to them as

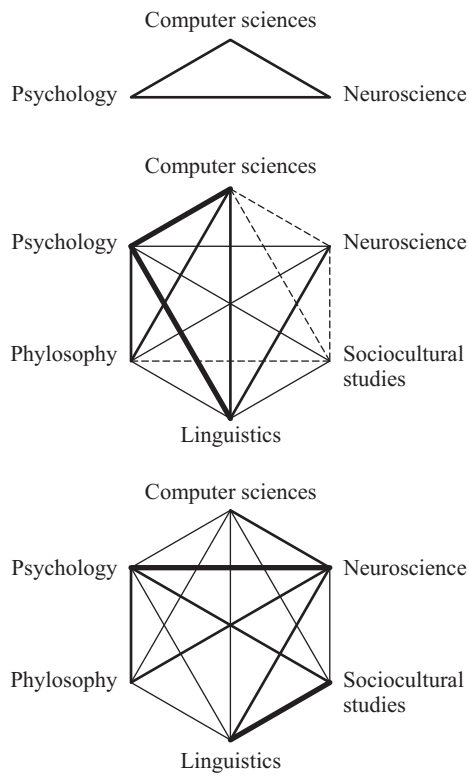


Figure 1. The history of cognitive science in three stages, from 1950 to 1998. The weight of the connections between and the size of the discipline labels denote relative importance of the respective disciplinary contributions for cognitive science (after Bechtel et al. 1998)

the standard (inside) history, or the "received view". The developments of the discipline seem to lend themselves to be grouped into three phases: formation and consolidation ("gestation"), followed by a phase of "maturation", and culminating in adulthood plagued by "identity crises" (Bechtel et al. 1998); see Figure 1.

Cognitive science began to be institutionalized in the United States in the late 1950s, early 1960s. This period is usually characterized by the emergence of its central conceptual framework (the computer metaphor, or computationalism), serving as a *lingua franca* for a variety of at least six disciplines (artificial intelligence, psychology, linguistics, neuroscience, philosophy, and later joined by anthropology). To exaggerate slightly, this energetic new approach modeled after the high technology of the day in one way or another succeeds in ending the supposedly near-absolute rule of radical behaviorism over US psychology, and makes it scientifically reputable again to study the higher cognitive functions of the human mind.

Interestingly, the perhaps most balanced treatment of this period of classical cognitive science is given by Howard Gardner, the chronicler, rather than historian, of the cognitive revolution, at least compared with the other standard accounts published much later. While Gardner is unfailingly supportive of early cognitive science, his support is not uncritical. He distills five key features characteristic of early cognitive science. Beyond the familiar allusion to the computer metaphor and interdisciplinarity, these include a space of autonomy given by the level of mental representation, and a strategic decision to disregard certain features of the mind such as emotion, background context, and historical or cultural factors (Gardner 1985: 6). These strategies will prove to be decisive for the subsequent development of the field, discussed below, and analyzed in the final section of Part one.

The second phase (Figure 1, middle) is roughly delineated by the emergence of rival conceptions of selected key features of cognitive science in, and between, several of the disciplines involved. These include debates about the nature of representation *(the imagery debate),* the modularity of cognitive functions such as language (attacked by *cognitive linguistics*), and most significantly, the nature of mental computation itself (the challenge raised by *parallel distributed processing* or *connectionism*). As the challengers (more or less) all argue from within the cognitive science establishment, this period has produced a flurry of activity in the philosophy of cognitive science: the conceptual foundations of the discipline have become hotly debated, with the contenders arguing back and forth about the validity of the concepts and research methodology in question. This is in clear contrast to the general lack of debate and analysis of the relationship between cognitive science and the research programs it succeeded and – according to the inside account – largely displaced.

The final phase (Figure 1, bottom) sees the emergence of an even greater variety of different takes on cognition and computation, with some new trends raising far more radical challenges to cognitive science. Unlike the first phase, this time the debate is not confined to selected key features of cognitive science. The new trends emphasizing the cultural environment *(situated cognition),* bodily action *(embodied action),* and the temporal and interactive nature of cognitive processes (*dynamical systems* approaches) now challenge the general orientation and framework of cognitive science, as well as its interdisciplinary model based on a *lingua franca* now thought to be too constraining. The last point is perhaps the most worrying, since even if the conceptual and methodological challenges can be met (as they were more or less successfully done in the previous phase), the sheer number of different multidisciplinary research programs now stretches a cognitive science worthy of the name (at least in the

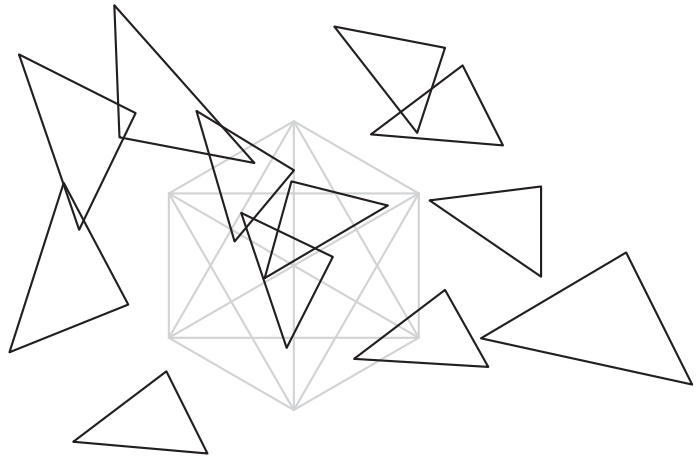


Figure 2. The "broken crystal": The current state of cognitive science is characterized by a growing sense of pluralism/fragmentation. The triangles denote multidisciplinary fields, some of them rather new, such as machine learning, cognitive ethology, computer-supported cooperative work, computational neuroscience, social neuroscience, human–computer interaction, etc. While many of these fields are at least somewhat sympathetic to the cognitive science enterprise – shown in the background as the familiar hexagon – these multidisciplinary fields now have their own conferences, journals, study programs, etc. In short, these fields project an increasing sense of "reality" compared to a cognitive science from which they benefit very little in terms of conceptual guidance, methodological integration, or social identity

singular) to its breaking point. "Cognitive science now recognizes some of the advantages overlooked in its development and has been drawn back downwards into the brain and outwards into the world [...] these and other factors will create tensions" according to Bechtel and his colleagues (1998: 98). As these tensions have continued in the last decade after their review, we think a more "graphic" illustration of the new state of pluralism or fragmentation is called for (see Figure 2).

Tacking Stock: Historical references and their uses during the three phases of cognitive science

The standard view of the history of cognitive science (Gardner 1985, Bechtel et al. 1998, Boden 2006) has at its core several elements strikingly similar to what is referred to as "Whig history" (Bowler and Morus 2005), which become apparent whenever recourse is made to events occurring before or close to the cognitive "revolution". This is to say that previous work is treated as a series of important precursors or stepping-stones to the great achievements of today – or alternatively discarded as hostile, at best a worthy enemy that has been overcome (see also Danziger 1997). As noted previously, the original account by Howard Gardner may actually be the most balanced and sophisticated of the prevailing historical treatises.[1] Gardner

[1] We will not discuss Margarete Boden's magnum opus in detail here, which for all its merits remains firmly entrenched in what we regard as internal debates in cognitive science.

even pays tribute to the history of (Western) ideas to the extent of including it among his five features of cognitive science and devotes half of his book to a discussion of the crucial developments laying the groundwork to the cognitive revolution within all of the six core disciplines of cognitive science.

We will discuss two related problems or deficiencies of the standard accounts here. First, even while the influence of behaviorism is clearly overplayed, it is never treated with anything remotely approaching the care such a supposedly momentous opponent would require (see Wozniak 1993). Instead, we get "radical behaviorism" dressed up as a straw man. Second, at the same time, cognitive scientists largely adopt the prevailing narrative of the larger history of psychology (going back to at least 1879 to the founding of the first psychological laboratory in Leipzig, Germany) from (behavioristic) US psychology, e.g., along the general lines of Boring (1950), even though the deficiencies of this account are well established by now (Danziger 1980). This is probably exacerbated by the difficulty in reading the (mainly German) literature of some of the then dominant schools of psychology before the center of gravity shifted to the English-speaking world. As a consequence, the radical differences in worldview and several fundamental shifts in the research methodology are downplayed or overlooked (Danziger 1990, 1997). This is probably a consequence of the inside history being mostly restricted to US history, or at best European developments directly influential in the US (which also helps explain the obsession with behaviorism, which never took root in Europe with the possible exception of Russian reflexology). This is understandable in the case of Howard Gardner given his focus on the early period; however, it is somewhat disconcerting to see Bechtel et al. (1998) practically ignoring Europe, Asia, and the rest of the world even at a time when there are substantial and active communities outside of the US.

The second phase of cognitive science (Figure 1, middle) is characterized by a very different use of history – with many of the proponents of the then new currents drawing heavily on research done right before or during the emergence of cognitive science. This "paradox" is noticed by Gardner in his epilogue to the second edition in 1987, and forms the backbone of Bechtel et al.'s account of this period characterized in terms of "re-discovering" the brain, neural networks, the environment, etc. Instances of this are easily found: artificial neural networks (Rumelhart, and McClelland 1987; building on McCulloch, and Pitts 1943), autonomous robots (Brooks 1991; taking up the explorations of William Grey Walter 1953), cognitive neuroscience (William Grey Walter again), computational neuroscience (von Neumann 1958, with a foreword by the Churchlands acknowledging the historical debts), differential equations and dynamical systems (Port, and van Gelder 1995; duplicating some of the work of Ashby 1952; see Grush 1997).

This is significant for cognitive science in that the innovative ideas are largely drawn from the past and are improved upon, in contrast to the original image of creating a new science of the mind perhaps inspired by some currents opposed to or preceding behaviorism.

As can be expected, this trend intensifies in the third period (Figure 1, bottom; see also Figure 2), and is still in force today. Significantly, we see several of the new trends creating their own histories, rather than adapting the dominant narratives of cognitive science. These "alternative histories" are broader still, as they are drawing from diverse scientific traditions as well as going further back in time.

A striking example is the (re-)introduction of the works of the Soviet psychologists (most famously Vygotsky, Leontiev, and Luria) to cognitive psychology, which gradually picks up

steam starting in the 1980s (see Cole 1998 for a discussion). Though dating back to the early 20th century (and to the "evil empire"), this radically different psychology has clearly resonated with many cognitive scientists dissatisfied with the focus on the level of individual cognition in psychology.

In recent years, proponents of situated cognition have refined the genealogy of their approach, and now offer systematic discussions of their antecedents. For example, Ed Hutchins traces his own approach of *cognitive ecology* backwards along three different lines: ecology of mind (Gregory Bateson), ecological psychology (J. J. Gibson), and activity theory (Leontiev) (Hutchins 2008), for an in-depth discussion see Clancey (2009), and Cole (1998).

Similarly, proponents of embodied action also devote a lot of time revisiting the work of a variety of previous scientists and scholars. It is not uncommon to see roboticists being inspired by the philosophy of Martin Heidegger, and the biology of von Uexkuell (Ziemke and Sharkey 2001), neurobiologists drawing on supposedly vanished traditions of ideo-motor theories (see A. Stock and C. Stock 2004, for a discussion), and to refer as far back as to Ernst Mach and Henri Poincaré (Rizzolatti, and Sinigaglia 2008), and even to create systematic historical treatises on the relationship between what we now refer to as neurobiology and psychology (Jeannerod 1985, 1996).

The important point here is that many of these lines of history have absolutely nothing to do with the ones selected to be included in the standard history of cognitive science.

An outside perspective on the history of cognitive science

So far, the focus has been squarely on cognitive science itself (while there are of course disagreements about what the extent and focus of cognitive science should be). However, since the recent developments draw heavily on a number of traditions which appear unrelated to cognitive science and go back in time a hundred years and more, it may be time for historians of cognitive science to rethink this strategy. We propose taking a step back from cognitive science, and re-locating the developments against the background of the larger history of scientific and scholarly traditions (such as psychology, philosophy, etc.). Here, we illustrate this approach by retracing the three phases of cognitive science from such a perspective. We shift our focus from cognitive science proper to its relationships with two selected traditions: biology, and the social sciences.

At several critical junctures, Howard Gardner lays the groundwork for such a perspective as an antidote to the inside history. Specifically, the original strategy of concentrating on rational thought rather than emotions, culture, and background context, is quite obviously incompatible with the interdisciplinary scope of cognitive science, at least as the term is usually understood. This denigration of emotions, background context, and socio-cultural factors can hardly be seen as a glowing invitation to join the "interdisciplinary" field – rather this strategy seems designed to exclude many schools and even whole disciplines such as behavioral biology, social psychology, not to mention psychoanalysis and cultural anthropology, the latter supposedly among the core disciplines of cognitive science. Jean-Pierre Dupuy follows this line of reasoning to its logical conclusion in his critical history of the cybernetic foundations of the cognitive revolution (Dupuy 1999). According to him, "cognitive philosophy", which forms the foundation of cognitive science, has "crept into the Trojan horse of these sciences

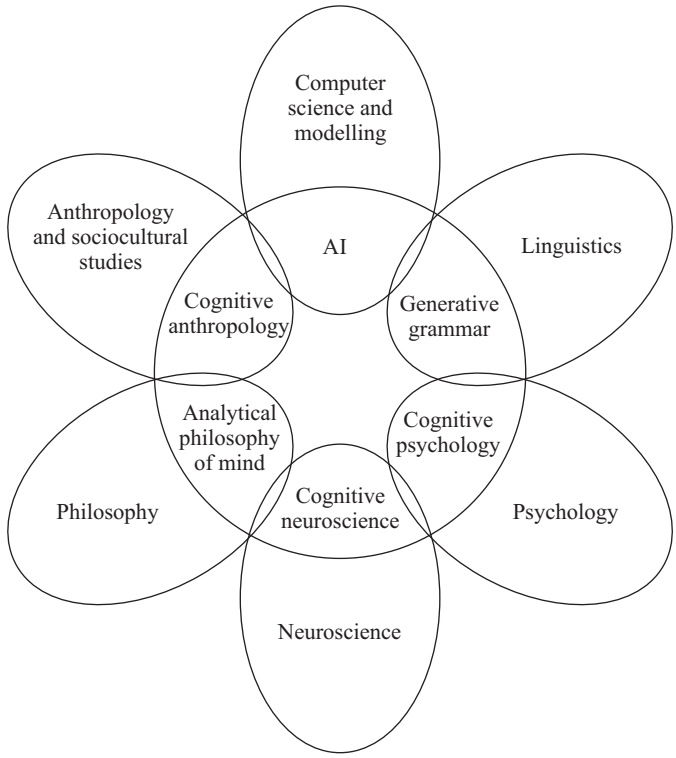


Figure 3. The interdisciplinary model of cognitive science viewed from an outside perspective. Rather than connecting the six core disciplines to each other (as implied by the familiar hexagon, see Figure 1) cognitive science is restricted to a selected few sub-fields or schools from the disciplines (depicted in the center and illustrated in terms of a clear-cut example). Most of the disciplinary research programs, concepts, and traditions remain firmly outside, and are indeed incompatible with, the cognitive science enterprise in its original formulation

[...] in order to assert dominion over the realm of the mind", and to "expel intruders", chiefly rival philosophies such as phenomenology, rival psychologies such as psychoanalysis or behaviorism, and rival sciences such as the social sciences (Dupuy 1999: 91). His argument is mainly about (analytic) philosophy but can be straightforwardly extended to all the disciplines involved in cognitive science: many, if not most, of the schools and research programs within the themselves rather diverse disciplines are actually excluded from (early) cognitive science. Instead, the interdisciplinary model admits only a selected subset of schools from the different disciplines, which have arguably more in common with each other than with the other subfields of their respective disciplines, at least with regard to subscribing to the computer metaphor of the mind (see Figure 3).

This coherence through self-restriction is in striking contrast to the earlier period when, in retrospect, the groundwork for cognitive science was laid. Surveying the contributors to the Hixon symposium and the Macy conferences, one is struck by the diversity of traditions, methodologies, and concerns included therein. There is not yet one clear "party line" that people have to adhere to – and, as a corollary, also no clear sense of institutional security or

academic identity apart from a consensus that behaviorism is too constraining. This severing from its cybernetic roots may have been the price that early cognitive science had to pay for "passing from an exploratory stage to a full-fledged research program – from a cloud to a crystal" in the words of Francisco Varela et al. (1991: 37).

From this perspective, the subsequent developments in cognitive science are essentially a steady expansion of cognitive science to include concepts, methodologies, and concerns more and more peripheral to its original center of gravity. Therefore, it is not surprising at all for the second phase of cognitive science to draw heavily on the time before the institutionalization of cognitive science to take up more and more of the lines of research pursued before the restriction to a manageable core (as discussed above).

These tendencies have had the effect of enlarging the scope of cognitive science to a point where the interdisciplinary rhetoric may finally begin to ring true. Ironically, cognitive scientists and philosophers more committed to defending the integrity of cognitive science than its lofty ideal of interdisciplinarity largely chose to fight a rearguard battle against these "attackers" before ultimately sanctioning the inclusion of these research programs into the discipline. Thus, the "mainstream" of cognitive science had to be expanded again and again, and with it its official history, culminating in the admission that cognitive science now "recognizes some of the advantages overlooked in its development" (Bechtel et al. 1998, quoted above), which serves to deflect attention away from the reasons for this strange exercise in tunnel-vision. After all, their set of mathematical and engineering tools (differential equations, artificial neural networks, autonomous robots, etc.) and an openness to physiology were central to cyberneticians, to whom all the official historians pay lip-service as their perhaps most important antecedents. How could most of their accomplishments have just been innocently overlooked or regarded as a waste of time simply because Marvin Minsky said so?

From the outside perspective, for all the apparent conflict and strenuously won expansion of cognitive science (and its official history), the developments of the second phase did not really constitute a crisis for the integrity of cognitive science. On the whole, the logic of the discipline as spelled out by Gardner, e.g., the predominance of artificial intelligence or the concept of cognitive processes as working on inputs to create outputs to the world, were largely left intact by the new research programs such as connectionism. Following Dupuy, one could characterize the different paradigms then co-existing within cognitive science as "members of a single extended, quarrelsome family" (1999: 90).

In contrast, the subsequent developments (latest phase, discussed above) really do constitute an ongoing crisis of cognitive science for two related reasons. First, the sheer number of research programs and multidisciplinary fields now relevant for cognitive science makes it more and more difficult to speak of (the) cognitive science(s) in any definitive way except as an umbrella term like the social sciences (which are infamous for their lack of coherence), as illustrated in Figure 2. Second, since the late 1980s, many influential proponents of the relatively new trends have openly attacked the very foundations of the cognitive science consensus. As alluded to above, the concept of representation independent of biology and society kept neuroscience (as well as other relevant parts of biology such as behavioral biology or human ecology) and the social sciences at arm's length from the start, even though they were supposedly among the core disciplines. From the outside perspective, the new trends of *embodied action* and *situated cognition* can be regarded as ongoing tendencies to overcome these self-imposed limitations of cognitive science by grounding cognitive processes

in physiology, and embedding them in their socio-cultural context, respectively. Indeed, it is very difficult to define these amorphous trends in any other way, as their "implications are all pervasive" (Clancey 2002), and they do not seem to have very much in common other than their shared commitment to "redress a perceived neglect" within cognitive science (Chrisley and Ziemke 2003) – which is eerily reminiscent of the early days before the so-called cognitive revolution, when reservations about the behaviorist orientation of psychology were perhaps the strongest bond between the participants of the Hixon symposium, for example.

Interestingly, Howard Gardner himself hints at the possibility of such struggles for the future of cognitive science. By placing its bets on the autonomy of the level of analysis (mental representation) kept "wholly separate from the biological or neurological, on the one hand, and the sociological or cultural, on the other", cognitive scientists may have created the conditions for consolidation to achieve a sense of integrity and identity for the fledgling discipline. Sooner or later, however, cognitive scientists "will have to discover or construct bridges connecting their discipline to neighboring areas of study", specifically "to neuroscience at the lower bound […] and to cultural studies at the upper". Otherwise "we will be left with a disembodied and incomplete discipline" (Gardner 1985: 44).

This "constructing of bridges" in the latest phase of cognitive science is reflected in the emergence of the alternative histories discussed above, due to the increasing number of cog-

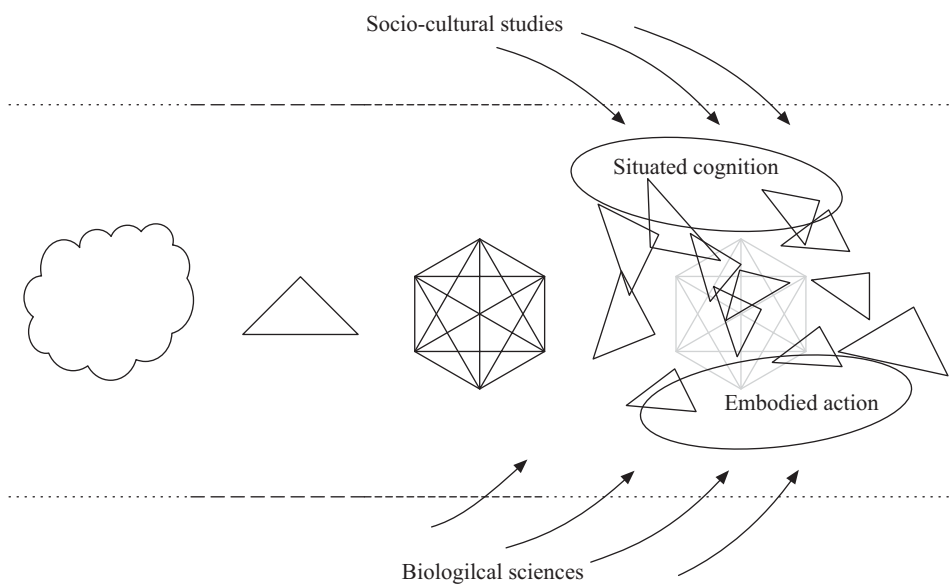


Figure 4. Cognitive Science between biology and the social sciences. The history of cognitive science is depicted here in the context of the relationship of cognitive science to biology and socio-cultural studies. After the institutionalization of cognitive science – its development "from a cloud to a crystal" –, openness to and interest in biology (chiefly neurobiology) and the social sciences (mainly anthropology) became highly restricted. From this outside perspective, much of the subsequent history of cognitive science can be regarded as a continuing process of the (re-)admission and (re-)integration of these traditions into cognitive science. This process culminates in the recent trends of embodied action and situated cognition, and contributes to the dissolution of the traditional model of the interdisciplinarity of cognitive science

nitive scientists now drawing on traditions explicitly (and not just temporarily) outside of cognitive science. The struggles for the conceptual foundations have clear parallels in the shifting academic identity of cognitive scientists: while part of the field has retained its clear-cut orientation as a behavioral science heavily relying on computer modeling, other parts would now self-identify as an applied science (see e.g., Rogers, Scaife, and Rizzo 2005), or as a biological, or as a neural and behavioral science (the cognitive neurosciences).

The history of cognitive science as seen from the outside perspective between biology and the social sciences is sketched in Figure 4.

Part two: The uses of history in cognitive science

> *[T]he actual history of a domain can only be constructed by articulating the differences of conflicting methodological–historiographic reconstructions. In no case will a single definitive or "true" history emerge from such reconstruction. (Weimer 1974)*

What are the agendas of the practitioner historians and what the functions of their histories?

The history of cognitive science was for the most part written by cognitive philosophers and scientists themselves, similarly to the history of (at least more recent) psychology (see Danziger 1990: vii).

The message of the inside historians (discussed above) is generally positive; their attitude towards their science is upbeat, in spite of, or even because of, the growing controversies: "we are optimistic that cognitive science will not only endure but develop into an even more interesting domain of science" (Bechtel et al. 1998: 98). The framing of the inside history reflects the current mainstream philosophy of cognitive science, which, while drawing heavily on the new trends, essentially minimizes their differences with traditional cognitive science (see Dennett 1993; Clark 1999). This parallel framing is hardly surprising as "[o]ur organization of the history of the field will also serve as a subtle justification of the way we have characterized the field in the present" (Danziger 1990: 1). Therefore, it is no exaggeration to conclude that the inside history of cognitive science, much like the inside history of psychology from which it takes its cue, ultimately serves to "cementing consensus among working scientists who can afford neither the disruptive effects of persistent controversy about fundamental issues nor the demoralizing effects, scepticism about the intellectual constructions on which their work is based" (Danziger 1997: 11).

As we have seen, the alternative histories of cognitive science are written by people who at least partly have different allegiances than mainstream cognitive science. To have one foot in a different field allows them to take a step back, and reflect on the relationship between cognitive science and one or more neighboring fields of study – thereby adopting and outside perspective on cognitive science.

Here we have focused on the work of Jean-Pierre Dupuy (1999), whose background is in analytic as well as continental philosophy, Francisco Varela whose background was in neurobiology and who has had a strong interest in Buddhist psychology and meditation practices (Varela et al. 1991), Ed Hutchins and Michael Cole (Hutchins 1996, 2008; Cole 1998) who have double background in cognitive psychology and cultural anthropology (plus an active

interest in Soviet psychology), and we have mentioned the work of Marc Jeannerod (1985, 1996), whose work combines the concerns of neurophysiology with cognitive psychology.

Unlike the inside historians whose agenda remains largely implicit, the advocacy of these inside-outsiders is on the whole more explicit, reflecting their commitment to "redress a perceived neglect" in cognitive science: by demonstrating how other research traditions intersect with the concerns of cognitive science, they buttress their case that their main area of interest (enaction, socio-cultural processes, the role of action in cognition, etc.) should be given more prominence in current cognitive science.

We conclude that, not surprisingly, the histories of cognitive science we have discussed here reflect the agendas of their authors, both in terms of the episodes and traditions selected, and the way the narrative is framed. The inside history serves to promote (the illusion of) the integrity of cognitive science and to protect the status quo, while the alternative histories set out to emphasize the historic differences between the concerns supposedly integrated in cognitive science and set out to change the field.

What is actually the proper subject matter of the history of cognitive science?

Can we not get rid of all this haggling and tinkering with history by turning to proper, objective, disinterested historiography? Obviously, any historiography must of necessity be highly selective in the material that it takes up and that it discards as irrelevant. Can "objective" criteria be found for this which do not depend on being familiar with the current developments in the field (which is hard to combine with a solid training in history), and partisan convictions about the future of cognitive science? Alas, the examples discussed in this paper would suggest otherwise:

We have seen that the very same research report dating back to the 1950s can be considered peripheral to cognitive science in the 1960s, "re-discovered" in the 1980s, and celebrated as a milestone in the 1990s (e.g., von Neumann 1958).

At first sight, this problem could be dealt with if we stick to reconstructing the actual lines of conceptual transmission (following Hull 1988), and record that this work was causally effective in cognitive science only much later after its publication (in effect, treating it as it was a paper published in the 1980s). However, on second thought, some people certainly were influenced by von Neumann's thought all along, and who can say that they really did not play much of a role in cognitive science before they became too successful to be ignored, and their favorite history was subsequently included in the standard account?

Even in case a historical work is (more or less) a genuine re-discovery (such as Vygotsky 1980), this will not get us off the hook. Why was it ignored for so long? Do the interests that chose not to include it still operate today? Moreover, the case of the "re-discovery" of Vygotsky's work illustrates that a historian who makes accessible, or otherwise promotes, "lost" historical episodes can have a significant impact on modern cognitive science.

Therefore, we conclude that the history of cognitive science is itself a moving target, being crucially dependent on – as well as possibly influencing – current events and agendas. The histories of cognitive science strikingly demonstrate that there is no room for neutrality.

This skepticism about neutrality emphasizes, rather than diminishes, the need for professional standards in historiography: we need to be very careful not to be held hostage to current

events, and resist the temptation to change our history to suit whoever is in power. Weimer (1974) strikingly demonstrates such retroactive cleansing of the historical record perpetrated by Boring. While Karl Bühler's work figured prominently in his first edition of his *History of experimental psychology,* Boring essentially stopped referencing his work (Boring 1950), thus reflecting Bühler's loss of power within the discipline as a consequence of his forced emigration. The historian is in a unique position to resist such Orwellian tendencies by insisting that we have not always been at war with Oceania.

What cognitive science can learn from (its) history

We have discussed how familiarity with current developments can help the historian to re-visit assumptions about the history of the field, and be more critical about the way his or her work may serve current agendas. As a corollary, a deeper understanding about the history of the research traditions that cognitive science draws on can help working cognitive scientists and students to get a clearer overview of their dynamic field, to better cope with conceptual inconsistencies, and to have a wider context for interpreting research questions and experimental results.

As discussed at the end of the first part, by looking at cognitive science within the context of the larger history of science, it is possible to discern patterns which the inside account misses outright or is forced to downplay (see Figure 4). Understanding these patterns can help cognitive scientists to navigate their increasingly complex field, and even cautiously predict the general outlines of future developments. Some of these developments would point to a still stronger set of interconnections between cognitive science, biology, and the social sciences, as recent trends show no sign of receding.

As the history of cognitive science strikingly illustrates, these interdisciplinarity relationships strongly depend on the dominant conceptions of the research methodologies suitable to address mental processes. These conceptions in turn partly reflect assumptions about the nature and workings of the mind which have a complex history in psychology. Studying this history can help to make these "background assumptions" more visible, as we are stepping "back a bit in time to a period when they were still controversial and therefore under active discussion" (Danziger 1997: 45).

To illustrate, we turn to recent work on the classic topic of reflex action. We now have systematic demonstrations of the "intelligence" of rapid motor responses (traditionally called reflexes), at many levels spanning from sensori-motor loops in the spinal chord (reviewed in Poppele and Bosco 2003) to the modifiability of these rapid motor responses by voluntary control (Pruszynski, Kurtzer, and Scott 2008). The basic setup and interpretation of the results crucially depend on novel frameworks derived from control theory (reviewed in Diedrichsen, Shadmehr, and Ivry 2009; Tresch and Jarc 2009). How will this research be received by "mainstream" cognitive science?

For a student of the history of the relationship between physiology and psychology, there is hardly a topic more central than the evolving concept of the reflex (a thorough discussion is given by Canguilhem 1975; Danziger 1983). However, after "importing" the concept from physiology, psychologists on the whole did not pay much attention to now classic research (e.g., by Anokhin in the 1960s) which has begun to sketch the physiological understanding of

the reflex in much more flexible and dynamic terms than the concept is laid out in psychology textbooks (reviewed in Berthoz 2002). More recently, classical cognitive science gave relatively short shrift to the work of Eric Kandel (1979). While his team was able to demonstrate the molecular mechanisms of simple adaptation and learning of conditioned reflexes, it did not cater to the prevailing concerns of cognitive science at the time: "habituation in *Aplysia* does not cry out for representational analysis". Instead, what was needed from neuroscience were "explanatory bridges between the level of the neuron and the level of the rule or the concept", the then dominant concepts (Gardner 1985: 287).

As illustrated in Figure 4, cognitive science has become much more open to physiological concepts since then, partly by adapting some of its conceptual core, such as the concept of representation. The prominence for cognitive science, or lack thereof, of the new research on rapid motor responses (e.g., Pruszynski, Kurtzer, and Scott 2008) will depend on how far these conceptual developments in cognitive science will go. The historical record would indicate that a rapprochement between the physiological, psychological, and control theoretic aspects of the reflex concept are possible, and indeed necessary to clear up the conceptual muddles brought about by conceptual change over centuries. After all, central cognitive concepts such as representation have been heavily influenced by research on the reflex done over a century ago (Danziger 1983), and the then state-of-the-art of control theory central to cybernetics (Dupuy 1999). These considerations would suggest that revisiting some of the roots of the related concepts of reflex response, action planning, and representation could address some of the inconsistencies that cognitive science has been battling since the 1990s (see also Grush 2004).

Conclusions

While there is a consensus about describing the brief history of cognitive science in terms of several phases, a trend towards expansion, and growing conceptual turmoil, the interpretation of these developments and even the kinds of historical sources relevant for cognitive science differ sharply among the practitioner historians. Not surprisingly, these differences reflect their agendas and concerns. The inside historians, whose focus is largely within cognitive science, tend to rework the narratives and include the (pre-)history of newly admitted programs (like the work on artificial neural networks by McCulloch and Pitts), while downplaying the shift in the orientation this entails. In contrast, the alternative historians, who have part of their background in a different field, construct their historical account to back up the contributions of these research traditions, and point out the gaps in cognitive science that need to be redressed.

Thus, the historical development of the relationship between cognitive science and its neighboring fields is itself an object of contention, and the accounts of their history both reflect and are intended to promote current agendas. It is against the background of historical development, both conceptual and methodological, that these relationships need to be analyzed. From this larger, outside perspective, it is possible to outline patterns in the history of cognitive science which reflect the degree of openness and interaction between the conceptual frameworks and research practices of cognitive scientists and biologists, on the one hand, and cognitive scientists and social scientists, on the other (see Figure 4). An appreciation of

the history of these turbulent developments may be crucial to better address the persistent problems resulting from basic differences between the "languages" of the biological, psychological, and the social sciences – and to assess the significance of current research in key areas, such as the reflex concept, for clarifying these contentious *boundary objects* between these fields.

References

Ashby, R. (1952) *Design for a Brain.* London: Chapman and Hall.

Bechtel, W., Graham, G., and Abrahamsen, A. (1998) The life of cognitive science. In: W. Bechtel, and G. Graham (eds.) *A Companion to Cognitive Science.* Blackwell Companions to Philosophy. Oxford: Blackwell, 1–105.

Berthoz, A. (2002) *The Brain's Sense of Movement.* London: Harvard University Press.

Boden, M. A. (2006) *Mind as Machine: A History of Cognitive Science.* New York: Oxford University Press.

Boring, E. G. (1950) *A History of Experimental Psychology.* 2nd ed. Englewood Cliffs (NJ): Prentice-Hall.

Bowler, P. J., and Morus, I. R. (2005) *Making Modern Science: A Historical Survey.* University of Chicago Press.

Brooks, R. A. (1991) Intelligence without representation. *Artificial Intelligence* 47, 139–159.

Canguilhem, G. (1975) Die Herausbildung des Reflexbegriffs im 17. und 18. Jahrhundert. München: Wilhelm Fink.

Chrisley, R., and Ziemke, T. (2003) Embodiment. In: L. Nadel (ed.) *Encyclopedia of Cognitive Science.* New York: Wiley & Sons, 1102–1108.

Clancey, W. J. (2009) Scientific antecedents of situated cognition. In: P. Robbins, and M. Aydede (eds.) *Cambridge Handbook of Situated Cognition.* New York: Cambridge University Press.

Clancey, W. J. (2002) Simulating activities: Relating motives, deliberation, and attentive coordination. *Cognitive Systems Research* 3 (3), 471–499.

Clark, A. (1999) An embodied cognitive science? *Trends in Cognitive Sciences* 3(9), 345–351.

Cole, M. (1998) *Cultural Psychology: A Once and Future Discipline.* Cambridge (Mass.): Belknap Press.

Danziger, K. (1980) The history of introspection reconsidered. *Journal of the History of the Behavioral Sciences* 16 (3).

Danziger, K. (1983) Origins of the schema of stimulated motion: Towards a pre-history of modern psychology. *History of Science* 21, 183–210.

Danziger, K. (1990) Constructing the subject: Historical origins of psychological research. New York: Cambridge University Press.

Danziger, K. (1997) Naming the Mind. How Psychology Found its Language. London: Sage.

Dennett, D. C. (1993) Review of F. Varela, E. Thompson and E. Rosch, The Embodied Mind. *American Journal of Psychology* 106, 121–6.

Diedrichsen, J., Shadmehr, R., and Ivry, R. B. (2009) The coordination of movement: Optimal feedback control and beyond. *Trends in Cognitive Sciences,* 31–39.

Dupuy, J. (1999) The Mechanization of the Mind: On the Origins of Cognitive Science. Princeton: Princeton University Press.

Gardner, H. (1985) The Mind's New Science: A History of the Cognitive Revolution. 2nd ed. New York: Basic Books.

Grush, R. (1997) Yet another design for a brain? Review of Port and van Gelder (eds.) *Mind as Motion. Philosophical Psychology* 10, 233–242.

Grush, R. (2004) The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and Brain Sciences* 27 (03), 377–396.

Hull, D. L. (1988) Science as a Process: An Evolutionary Account of the Social and Conceptual Development of Science. Chicago (IL): University of Chicago Press.

Hutchins, E. (2008) Cognitive ecology. (Presented at the 30th Anniversary Symposium: *Trajectories of Cognitive Science.*) Washington DC.

Hutchins, E. (1996) *Cognition in the Wild.* Cambridge (MA): MIT Press.

Jeannerod, M. (1985) The Brain Machine. The Development of Neurophysiological Thought. Cambridge (MA): Harvard University Press.

Jeannerod, M. (1996) De la Physiologie Mentale: Histoire des Relations entre Biologie et Psychologie. Paris: Odile Jacob.

Kandel, E. R. (1979) Small systems of neurons. *Scientific American* 241 (3), 60–70.

McCulloch, W. S., and Pitts, W. H. (1943) A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics* 5, 115–133.

von Neumann, J. (1958) The Computer and the Brain. With a foreword by Paul M. Churchland and Patricia S. Churchland. Yale: Yale University Press.

Poppele, R., and Bosco, G. (2003) Sophisticated spinal contributions to motor control. *TRENDS in Neurosciences* 26 (5), 269–276.

Port, R., and van Gelder, T. J. (1995) *Mind as Motion: Explorations in the Dynamics of Cognition.* Cambridge (MA): MIT Press.

Pruszynski, J. A., Kurtzer, I., and Scott, S. H. (2008) Rapid motor responses are appropriately tuned to the metrics of a visuospatial task. *Journal of Neurophysiology* 100 (1), 224.

Rizzolatti, G., and Sinigaglia, C. (2008) *Mirrors in the Brain: How Our Minds Share Actions and Emotions.* New York: Oxford University Press.

Rogers, Y., Scaife, M., and Rizzo, A. (2005) Interdisciplinarity: An emergent or engineered process? In: S. J. Derry, C. D. Schunn, and M. A. Gernsbacher (eds.) *Interdisciplinary Collaboration: An Emerging Cognitive Science*. Mahwah (NJ): Lawrence Erlbaum, 265–286.

Rumelhart, D. E., and McClelland, J. L. (1987) Parallel Distributed Processing, Exploration in the Microstructure of Cognition. Vol. 1. Foundations. Cambridge (MA): MIT Press.

Stock, A., and Stock, C. (2004) A short history of ideo-motor action. *Psychological Research* 68 (2), 176–188.

Tresch, M. C., and Jarc, A. (2009) The case for and against muscle synergies. *Current Opinion in Neurobiology* 9 (6), 601–607.

Varela, F. J., Thompson, E., and Rosch, E. (1991) *The Embodied Mind: Cognitive Science and Human Experience.* Cambridge (MA): MIT Press.

Vygotsky, L. S. (1980) *Mind in society.* Cambridge (MA): Harvard University Press.

Walter, W. G. (1953) *The Living Brain.* New York: Norton.

Weimer, W. B. (1974) The history of psychology and its retrieval from historiography. Vol. I. The problematic nature of history. *Science Studies* 4, 235–258.

Wozniak, R. H. (1993) Theoretical Roots of Early Behaviourism: Functionalism, the Critique of Introspection, and the Nature and Evolution of Consciousness. History of Psychology: The Roots of Behaviourism. London: Routledge.

Ziemke, T., and Sharkey, N. E. (2001) A stroll through the worlds of robots and animals: Applying Jakob von Uexkülls theory of meaning to adaptive robots and artificial life. *Semiotica* 2001 (134), 701–746.

MIND THE HISTORICAL GAPS: COGNITIVE SCIENCE AND ECONOMICS[1]

Filomena de Sousa

Introduction

Literature on mind is often unmindful: oblivious of economics as piece and parcel of the interdisciplinary effort that gave rise to cognitive science as an independent branch of science.

This is a curious gap for interpretations of how the human mind works, playing a central role for economists.[2] The interface between economics and disciplines devoted to the study of cognitive processes was shaped by mutual imports and exports.[3] So much so that two Nobel Laureates in economics were the founding fathers of models that shaped cognitive science, Cognitivism and Symbolism. Hayek and Simon's contributions played at several levels illustrating integration of different competing strands and disciplinary approaches that gave rise to cognitive science.[4] Their role was even more so significant for despite a broad cross-disciplinary integration, artificial intelligence, pioneered by Simon, and psychology, where Hayek innovated, stand as locus of the enterprise out of which cognitive science was born. My purpose is to outline their place in the history of cognitive science with contributions evidencing contrasts and convergence between Connectionism and Symbolism.

However superseded by recent research, Hayek and Simon offered resources whose consequences remain relevant for current debates. Their shared concerns were to understand how the mind works. For Hayek, this involved understanding how physiological processes are structured and relate to consciousness. For Simon, the strategy was psychotechnics, understanding how one can build a mind. Their lasting importance in the field of cognitive science is reflected in the challenging questions they raised, which are still waiting for a suitable solution.

Hayek: Precursor of connectionism

Literature on cognitive science largely neglects Hayek's *The Sensory Order* (henceforth SO), drafted in the 1920s, reviewed and published in 1952, as one of the first comprehensive

[1] I am indebted to the Foundation for Science and Technology of Portugal for financial support, and to the University of Cambridge for welcoming a stimulating intellectual atmosphere during my research leave. To Gonzalo Munevar goes my gratitude for inspirational comments. The usual caveat applies.

[2] The interface between economics and cognitive science has widened since the 1990s, opening research fields like neuroeconomics, e.g., Glimcher (2004); Ross (2005); Frey and Stutzer (2007). For a challenging account, see Mirowski (2002).

[3] See Ross (2005) on economics as a self-contained versus massively integrated with behavioural and cognitive sciences.

[4] See Gardner (1985) for a history of the cognitive revolution.

connectionist models. Yet the model of mind it offers, a self-organising system whose order arises out of decentralised units is a crucial source of neural network models (Birner 1994: 1–21).

Hayek started to be credited as the precursor of connectionism only in the course of the 2000s. In Mirowski's gallery of figures that dragged economics into the field of cyborg sciences, Hayek is dubbed as a "pivotal agent provocateur in disseminating the germs of these cyborg themes" (Mirowski 2002: 235) Convergence between cognitive and economic arguments characterises Austrian economics: socioeconomic institutions perform the role of a collective mind but the problem is explaining how it works. Hayek's whole *corpus* centring on dispersed information and spontaneous coordination in a context of complexity is clearly underwritten by the theory of mind he devised while a student.[5]

Towards the end of his career, Hayek avowed he "chose economics, perhaps wrongly; the fascination of physiological psychology never quite left me" (Hayek 1982: 288). He estimated SO as his most important contribution to knowledge, and insights gained while researching the essay as the most exciting intellectual events that marked his thinking (Hayek 1994: 153).[6]

Hayek took up issues that were virtually unsolvable in the context of the 1920s' psychology (Birner 1994: 1–21). He described himself as a 19th-century ghost without access to a psychology professor, self-taught through sources in line with his interest in Weber,[7] namely, Müller, Helmholtz and Wundt (Hayek 1982: 287), but first and foremost Mach (Hayek 1952: vi). From the founders of experimental psychology, Hayek assimilated the insight that the mind viewed as substance was untenable, and that consciousness is an activity consisting of perpetually interacting processes. His proposals, nonetheless, deviated from his intellectual ancestors. His interpretation of Mach was riddled with ambiguities, as Hayek staunchly opposed positivism and shunned from neutral monism. Yet Mach's theory of perceptual organisation provided much of SO infrastructure.[8]

Notwithstanding dissenting from behaviourism, Hayek outlined points of convergence, particularly with Ryle. Gestalt psychology made significant inroads in Hayek's *corpus* with frequent references to Kohler and Koffka. Hayek was also familiar with Dewey and James. Pragmatism and functionalism shaped much of his theoretical psychology.

During his LSE times, he got acquainted with British psychology, and upon moving to Chicago, Hayek familiarised himself with modern complexity theory.[9] He drew from scholars hailing from diverse disciplinary fields that contributed to the gestation of cognitive science: Ross Ashby, W. McDougall, Eccles, Sherrington, D. Campbell, Lashley, Chomsky, and Piaget, among others.[10]

[5] Rizzello (1999) offers a sophisticated analysis of Hayek and Simon's research on economics and mind.

[6] For an appraisal of *The Sensory Order* within the context of Hayek's work, see Gray (1984); Birner (1994); de Sousa (2005).

[7] For Weber's interaction with Austrian economists, namely Mises, see de Sousa (2005). Hayek's plans to study under Weber were foiled by the latter's death.

[8] See de Vries (1994) for Hayek–Mach affinities and divergence.

[9] For Hayek's intellectual background, see Caldwell (2004).

[10] Hayek mentions Külpe, but surprisingly omits Würzburg psychologists whose views converged with his, namely Selz, and Karl Bühler, who introduced Kantian ideas in Austria.

However, throughout his career, Hayek held foremost to his Viennese background. His work on the mind converges with ideas of fellow Austrian expatriates and American collaborators that chipped in stepping-stones for cognitive science: Wiener, von Foerster, McCulloch, and Polanyi. But particularly, his friends Popper and von Bertalanffy who read and commented on a draft of SO. Hayek became acquainted with Hebb's ideas only after SO's completion, and wondered if his work was still worth publishing. He decided their works were complementary.

SO's chief purpose was not to elucidate how the mind works, but how it is embodied: Hayek departed from Mach's neutral monism framing the problem of perception in terms of mind–body problem. Notwithstanding straddling diverse frameworks, his position can be pigeonholed as a form of functionalism (de Vries 1994) accommodating non-reductive physicalism (Birner 1994). Hayek's world is essentially a layered one committed to some form of supervenience.

Hayek articulates his vision around three orders. At the onset, he distinguishes the phenomenal or sensory order from the physical world: the "subjective, sensory, sensible, perceptual, familiar, behavioural, or phenomenal world", and the "objective, scientific, 'geographical', physical, or sometimes 'constructional'" (Hayek 1952: 1.11).

Hayek proceeds with a dichotomy between the order of "different attributes or dimensions with regard to which we differentiate in our responses to different stimuli", that is the cerebral cortex, and consciousness or the subjective order, nowadays corresponding to the qualia problem. The subjective order is described as the realm of "affective qualities and the mental 'values' which make up the more comprehensive order of 'mental qualities'" (Hayek 1952: 1.5). The three orders, the outer or physical world; the sensory cortex which in a sense is part of the former order but which Hayek differentiates as the sensory order; and the subjective[11] order of consciousness which, in turn, is a sub-set of the former.

Orders are purely relational. The problem, however, is how orders relate to one another. Hayek's answer comes in the currency of cognitive associative architecture based on activation that flows through a network of links. Mental properties are conceived as relational features of events in the sub-symbolic realm of nerve excitations. Different neural networks are present in Hayek's work, but his picture of cognition comes closer to the predominant version, parallel distributed processing: mental processing involves dynamic graded evolution in a neural net, the activation of each unit depending on connection strengths and the activity of its neighbours according to the activation function.

Sense impressions and mental phenomena acquire meaning by virtue of their location in a structure of myriad of neural connections, a spontaneous order of sensory qualities "even though to us, whose mind is the totality of the relations constituting that order, it may not appear as such" (Hayek 1952: 1.56). Hayek objected to isomorphism the "particular misunderstanding of the idea of a one-to-one correspondence between impulse and sensation" (Hayek 1952: 2.13). He specifically criticizes J. Müller's theory and reformulations by Boring, E. Hering, Weiss, and Sperry.

SO's central argument is the recognition "that the differentiation of the sensory qualities is not due to the communication [it] does by no means make the conclusion inevitable that it must then be a difference in the properties of the impulses taking place in the different fibres,

[11] Austrian economics, whose hallmark is subjectivism, centres on this order.

which accounts for them [… The] specific character of the effect of a particular impulse need be neither due to the attributes of the stimulus which caused it, nor to the attributes of the stimulus which caused it, but may be determined by the position in the structure of the nervous system of the fibre which carries the impulse" (Hayek 1952: 1.35).

To perceive an event is to place it in a relational web processed by our nervous system through a classification apparatus involving a "map" of semi-permanent neural connections resulting from past experiences, "a construction set which supplies the parts from which the models of particular situations can be built" (Hayek 1952: 5.89). Within this topology, "relevant relations between the individual points are not their spatial relations, however, but solely the paths through which impulses can be transmitted" (Hayek 1952: 5.29).

New inputs force the human mind to reclassify phenomenal reality leading to structural change. Models refer to organism and environment interaction, more fluid than the map; they are closely related to learning. The sensory structure or map represents only the world part which the organism had previous sensory experienced and is subject to "continuous modification by new linkages between impulses" (Hayek 1952: 5.14). Cognition as pattern-mapping was given diverse formulations in modern connectionist models as the hallmark of neural networks and PDP.[12] It fits in well with externalism, depicting perception as a highly context-dependent process.

In contemporary connectionism, cortical memory networks are created through strength shifting in neural units connection. Memory is active competence, a system of processing patterns subject to change, constantly and cumulatively affected by the processing of past experiences for there is no "physiological mechanism which can retain anything except connexions between different events" (Hayek 1952: 5.12).

Hayek objected to memory-storehouse, the "conception that with every experience some new mental entity representing sensations or images enters the mind or the brain and is there retained until it is returned at the appropriate moment" (Hayek 1952: 5.11). This, he believed, was an offspring of "the theory of the absolute character of sensory qualities, and connected with the erroneous interpretation of the theory of the specific theory of the nerves" (Hayek 1952: 5.12).

SO's central thesis is that we "do not first have sensations which are then preserved by memory, but it is a result of physiological memory that the physiological impulses are converted into sensations. Connexions between the physiological elements are the primary phenomenon which create the mental phenomena" (Hayek 1952: 2.50). This raises the vexing question of nativism that Hayek tried to elude.

The dilemma was a consequence of Hayek's grappling with the terminology and interpretation of the two positions as involving twofold problems. The first is whether "the order of sensory qualities is congenital or acquired by individual experience. On this probably no general answer is possible" (Hayek 1952: 5.15). The second is whether the sensory order "is based on the retention of connexions between effects exercised upon them by the external world. With regard to this second question our answer is definitively empiricist" (Hayek 1952: 5.15).

Emphasis on the mind as a constructive system and active mental representation fostered Hayek's suspicion of nativism. The outcome was a view of the mind as *tabula rasa* suggested

[12] Bechtel and Abrahamsen (1991) are emblematic exponents.

in the claim that "the theory developed here traces all sensory qualities, 'elementary' as well as gestalt qualities, to the pre-sensory formation of a network of connexions based on linkages between non-mental elements" (Hayek 1952: 5.16).

Empiricism is hardly reconcilable with Hayek's framework and Kantianism running through Austrian economics.[13] Kantian scepticism is evidenced in SO, as well as the idea that experience "does not begin with sensations or perceptions, but necessarily precedes them: it operates on physiological events and arranges them into a structure or order which becomes the basis of their 'mental' significance" (Hayek 1952: 8.5). We need to bear in mind, however, that Kant's formal rules converge with symbolism, whilst Hayek posited tacit regulative principles. Moreover, Hayek's patterns of classification defining sensations emerge in the process of perception, and are not as fixed or *a priori* as in Kant.

Chomsky's linguistics exposing the failure of stimulus account caught Hayek's attention. Both concurred that rules of cognition are not a permanent feature of the mind's physical structure so that continuing transformation of cognitive categories begs an evolutionary explanation. SO's fuzzy physiological infrastructure resulting from evolution but activated by experience was reinforced in reference to Chomsky in later works[14] despite Hayek's opacity as to how much of the sensory order is species hardwired, and what results from individual experience.

The distributed manner in which connectionist systems store information raises a problem in relation to formal reasoning, for they seem to lack structures for identifying propositional attitudes and abstract reasoning. This poses a problem for neoclassical economics. Austrians, however, were concerned with the tacit knowledge that does not need be explicitly represented. One of SO's merits was precisely the way it explained mental events that are not conscious, dispelling previous common definitional fusions of mental with conscious (Weimer 1982: 283).

Notwithstanding his focus on basic cognitive processes, Hayek did not neglect symbolic processing. He assumed a hierarchical system rather than a highly distributed model. The principle of classification of elementary sensory qualities "applies also to the so-called 'higher' mental processes such as the formation of abstract concepts and conceptual thought" (Hayek 1952: 3.77). This oversimplified explanation cannot be taken but as a principled stance. Nonetheless, one might infer that in the same way as he reconciled Chomsky's model with his connectionist view of mind, Hayek would probably have sided with modern connectionists that seek to implement a symbolic processor within neural networks to describe formal thought.[15]

More questionable, but consistent with Hayek's overall views on complexity are the consequences this model of cognition has for science. Complexity and the limits of knowledge, keystones of Hayek's corpus implied that "mental or phenomenal order of sensations (and other mental qualities) [are] directly known although our knowledge of it is largely only a

[13] Hayek shared Piaget's premise that knowledge continues biological adaptation by different means, a point on which he gave a more consistent naturalist articulation in later works.

[14] See Hayek (1967, 1973).

[15] For example, Smolensky et al. (1992).

'knowing how' and not a 'knowing that', and although we may never be able to bring out by analysis all the relations which determine that order" (Hayek 1952: 2.7).[16]

If perception is classification/reclassification on successive levels, we can never perceive unique properties of individual objects but of properties they have in common with other objects, even if we are unaware of this fact. It follows that explanation "is always generic in the sense that it always refers to features which are common to all phenomena of a certain kind, and it can never explain everything to be observed on a particular set of events" (Hayek 1952: 8.57).

Explanations are achieved by generalization, and complex phenomena confined to explanations of the principle, or generalisations. With the human brain as classification apparatus, although "we may understand its modus operandi in general terms, or, in other words, possess an explanation of the principle on which it operates, we shall never, by means of the same brain, be able to arrive at a detailed explanation of its workings" (Hayek 1952: 8.80).

Hayek dooms us to the ignorance of mental processes' inner structure as a consequence of the principled stance predicated upon his belief that "any apparatus of classification must be of higher degree of complexity than what it classifies, and that therefore the human brain can never fully explain itself, may be inadequate" (Hayek 1982: 292) – a principle consistently invoked in relation to artificial intelligence. Knowledge about and derived from whichever mechanism we build is limited by "the fact that any apparatus of classification must be of a higher degree of complexity than what it classifies" (Hayek 1982: 292).

After the early 1960s, evolutionary theory becomes the theoretical framework for Hayekian spontaneous orders and a reformulation of the mind–body problem: the mind belongs to the realm of biology. It was SO that paved the way for the evolutionary stance with an awareness that an "adequate account of the action of the central nervous system would require as its foundation a more generally accepted biological theory of the nature of adaptive and purposive processes than is yet available" (Hayek 1952: 4.5). As Hayek confided, "conception of evolution, of a spontaneous order and of the methods and limits of our endeavours to explain complex phenomena have been formed largely in the course of the work on that book" (Hayek 1979: 199).

Simon: Builder of information storage systems

Simon's disapproval of public administration and economics disconnected from reality set him on the route to pioneer two scientific disciplines, cognitive psychology and artificial intelligence. Hayek's stumbling blocks played a reversed role in Simon's program spurring research on hierarchic systems and decomposability from a computational perspective.[17] Simon's approach to complexity is of an altogether different sort than Hayek's. He approached

[16] Hayek's reference to Ryle.

[17] What joins Hayek and Simon is also what sets them apart from economics Nobel Laureates close to their generation, Paul Samuelson and Kenneth Arrow concerned with logical implications and the formal expression of neoclassical rationality.

cognition from an engineering angle: Humans cannot master but can handle middle-range complexity.

The quest that defined his career can be traced to Administrative Behaviour (1947), his doctoral dissertation addressing behavioural and cognitive processes underlying choices that challenged the neoclassical conception of substantive rationality.[18] There he articulates notions of causality, near-decomposability in complex systems, and operational definitions of political organizational power as pillars for his theory of decision-making in conditions of uncertainty and interdependence as key factors to rebuild an organisation's internal mechanisms. The book, he later stressed, contained the "superstructure of the theory of bounded rationality that has been my lodestar for nearly fifty years" (Simon 1991: 86).

The first ideas he assembled came from similar resources that Hayek drew upon, Gestalt psychology and Wundt's insights on cognition and learning as processes of adaptation to the environment.[19] Human thinking is confined to proximate solutions to mental problems that themselves only approximate real-world problems. Decision-making in conditions of uncertainty and interdependence, thus, begged mechanisms to compensate for limited rationality. This was the idea he explored throughout his career: knowledge as context-bounded problem solving.

The knowledge of economic tools and the mainstream concept of rationality that he set out to discredit came from a source that also steered his work in the direction of artificial intelligence, involvement with the Cowles Commission[20] whose greatest impact "was to encourage me to try to mathematize my previous research in organization theory and decision making" (Simon 1947: 4).

Simon grounded his proposals on procedural or instrumental reason: operating factually on value input, and serially on one problem at a time. As the overcome of cognitive limitations, Simon articulated in *Administrative Behaviour* a decision-making scheme, chains of decision guided by hierarchical means ends. His work on organisations suggested that humans rely on stock recipes, or heuristics, rather than seeking optimal solution procedures. As restated throughout his *corpus,* to "escape from the difficulty that, in a complex world, the alternatives of action are not given but must be sought" (Simon 1979: 3), the solution is trading "optimizing" for "satisfying", a problem-solving searching mechanism cutting down on decision time in a complex world. The strategy is finding solutions "good enough" employing rules of thumb or heuristics to facilitate decisions in a less than perfect environment illustrate "bounded rationality". This insight, guided his "whole scientific output" (Simon 1991: 88).

Simon's research strategy rested on the view that the mind does, in part, perform cognitive tasks by computing. Computation is symbol-processing, any intelligent system, mind or computer, manipulates information represented by strings of symbols. At Cowles meetings, papers discussed featured von Neumann's print for digital computer architecture, with data and programme stored together, which stimulated Simon's interest in the computer simulation of human cognition. The best way to study problem-solving was to simulate it through

[18] Simon was affected by the doctrine that dominated the University of Chicago in 1930. From his experience with Ridley on the municipal reform movement, he concluded that existing administration principles were vague and contradictory, useless practical guides to be consigned to a Carnapian hell.

[19] Hayek (1952: 7.13).

[20] See Augier and March (2004: 3–32).

computer programmes. Neurological associationalism and Gestalt explanations took the form of symbolic information processing within the frame of computational cognitive modelling.

Computers transported "symbol systems from the platonic heaven of ideas to the empirical world of actual processes carried out by machines or brains" (Simon 1996: 22–23). The turning point came in 1956 with Simon and Newell's implementation of the first artificial intelligence programme, Logical Theorist, on the Rand Jonniac.[21] The requirement for complex cognition is that the system, brain or computer, be capable of processing information represented in data-structures or symbols to which rules apply. Memory as a hierarchical store-box follows from these premises. Simon remained stern regarding the "claim that the human cognitive system is basically serial [which] has been challenged in recent years by advocates of neural nets and parallel connectionist models of the nervous system" (Simon 1996: 81).

High symbol manipulation encompasses human and artificial intelligence, although emphasis on trial-and-error search processes guided by heuristics set Newell and Simon's work apart from more algorithmic models in the field. The strength of Simon's strategy was to "avoid explicit consideration of the formal theory of computation and instead to build computer simulations of economic and mental phenomena, largely avoiding prior neoclassical models [… There] is only a certain 'middle-range' of observed phenomena of a particular complexity that it is even possible for us mere mortals to understand; and since reality is modular, we might as well simulate these accessible discrete sub-systems" (Mirowski 2002: 529). Simon saw the computer as the paradigmatic simulation machine both capturing our limitations and indicating more efficient means of behaviour.

This does not mean that the computer is a metaphor for mind, or that architectures of modern digital computers can give us insights into human mental architecture. Simon's programme rests on some level of analogy between human mind operations and computer software. But Simon and Newell also resisted a strict analogy and did not assume a structural identity between logical and mental operations. Crowther-Heyck's contention that the "mind–body problem was solved by analogy between mind and programme, computer and body" (Crowther-Heyck 2005: 274) is misleading. Simon was very cautious on this point. The mind is located in the brain but his research focused on "the organization of the mind without saying anything about the structure of the brain" (Simon 1996: 83). He explained the mind's disembodiment as an expository strategy to express the frontier between neurophysiology and information processing that can be realised in the human brain, the computer hardware, or eventually in other systems (Simon 1979: 1996).

Discrepancy between computer and brain creeps in when it comes to tacit knowledge and cognition/affect interaction (Simon 1979: 29–38). In his heuristics model, Simon defined satisfying as a searching "stopping rule" triggered by emotion when a good enough solution to a problem is found. Faced with Neisser's objections, Simon acknowledged that information processing theories needed to take into account the "intimate association of cognitive processes with emotions and feelings, and determination of behaviour by the operation of a multiplicity of motivations operating simultaneously" (Simon 1979: 38). His reply to Neisser captures the whole meaning of his enterprise, to "discuss the behaviour of humans, not the capabilities of computers. Nonetheless, Neisser has characterised correctly

[21] See Feigenbaum's (2004) narrative on Simon announcing to his students that he and Newell invented a "thinking machine".

some of the visibly gross differences between human behaviour and the behaviour of existing simulation programs" (Simon 1979: 29).

Simon was aware that intuition was a crucial decision-making mechanisms linked with emotion. Intuition exploits knowledge gained through past searches, thereby, cooperating in problem-solving (Simon 1983: 23–29).[22] Recent resurgence of interest in his work owes much to research on cognitive tools linked to emotions, particularly works by intellectual heirs Gigerenzer and Selten (2001).

Similar tension arises in connection with evolutionary analogy. A natural link flew from his work on complexity based on bounded rationality to sciences of the artificial and evolutionary theory. From servo-mechanisms he assumed functional equivalence between organisms, organisations, and machines. Feedback was an essential component of adaptive systems, organic and mechanical, that could evolve complex behaviours by nesting (Crowther-Heyck 2005: 319). Bounded rationality focuses on processes underlying judgement and choice requiring selective search to discover adaptive responses like in Darwinian evolution. The generator-test is a direct analogue, in the behavioural theory of rationality, of the Darwinian variation–selection mechanism (Simon 1983: 40–41). However, in contrast with biological evolution, instrumental rationality allows for engineering within a limited look-ahead framework (Simon 1983: 66).

Bridges leading to a destination

Hayek was silent regarding Simon who overtly outlined shared assumptions underlying their work: emphasis on the limited scope of human rationality in a context of complexity, an evolutionary perspective of mind and learning.[23]

But whereas Hayek stressed insurmountable human limits on the face of complexity, and the divide between natural and social sciences, Simon held that the "task of a natural science is to make the wonderful commonplace: to sow that complexity, correctly viewed, is only a mask for simplicity; to find pattern hidden in apparent chaos" (Simon 1996: 1). His whole efforts were directed towards an articulation of procedures to palliate reason's bounds, for he ultimately trusted a theoretically robust and practically useful science of human behaviour.

Hayek and Simon's views on mind show that neither symbolism, nor connectionism are free of liability. Neural networks are appealing descriptions of human cognition for they can modify their structure. Artificial networks are rigid interlinking lines or strings of information.

Much of the debate over cognitive architecture since the birth of cognitive science has centred on the alternative views expressed in Hayek and Simon's work. Much of later research where adaptation and active perception take centre stage, as in Hayek and Simon's work, albeit in different ways, tends to show that the gulf between the two models is not as wide as initially thought and that the two approaches need not be at odds.[24] Bridge building

[22] Simon's vision of intuition bears affinities with tacit knowledge prominent in Ryle, Polanyi and Hayek.
[23] See particularly Simon (1983).
[24] Smolensky et al. (1992), and Rowlands (1999) illustrate nicely the attempts of integration.

depends as much on the theoretical reappraisal of principled assumptions as on future empirical research.

References

Augier, M., and March, J. (2004) Herbert A. Simon, Scientist. In: M. Augier, and J. March (eds.) *Models of a Man: Essays in Memory of Herbert A. Simon.* Cambridge (MA): MIT Press, 3–32.

Bechtel, W., and Abrahamsen, A. (1991) *Connectionism and the Mind.* Cambridge (MA): Basil Blackwell.

Birner, J. (1994) Introduction: Hayek's grand research programme. In: J. Birner, and R. van Zijp (eds.) *Hayek: Co-ordination and Evolution.* London: Routledge, 1–21.

Caldwell, B. (2004) Hayek's Challenge: An Intellectual Biography of F. A. Hayek. Chicago: University of Chicago Press.

Crowther-Heyck, H. (2005) *Herbert A. Simon: The Bounds of Reason in Modern America.* Baltimore: Johns Hopkins University Press.

Feigenbaum, E. A. (2004) On a Different Plane. In: M. Augier, and J. March (eds.) *Models of a Man: Essays in Memory of Herbert A. Simon.* Cambridge (MA): MIT Press, 383–388.

Frey, B. S., and Stutzer, A. (eds.) (2007) *Economics and Psychology.* Cambridge (MA): MIT Press.

Gardner, H. (1985) *The Mind's New Science.* New York: Basic Books.

Gigerenzer, G., and Selten, R. (2001) *Bounded Rationality and the Adaptive Toolbox.* Cambridge (MA): MIT Press.

Glimcher, P. W. (2004) Decisions, Uncertainty, and the Brain: The Science of Neuroeconomics. Cambridge (MA): MIT Press.

Gray, J. (1984) *Hayek on Liberty.* Oxford: Basil Blackwell.

Hayek, F. A. (1952) The Sensory Order: An Inquiry into the Foundations of Theoretical Psychology. London: Routledge & Kegan Paul.

Hayek, F. A. (1967) *Studies in Philosophy, Politics and Economics.* London: Routledge & Kegan Paul.

Hayek, F. A. (1979) *Law, Legislation and Liberty.* Vol. III. London: Routledge & Kegan Paul.

Hayek, F. A. (1982) The Sensory Order After 25 years. In: Weimer, W. B., and D. S. Palermo (eds.) *Cognition and the Symbolic Process.* Vol. 2. Hillsdale (NJ): Lawrence Erlbaum, 241–285.

Hayek, F. A. (1994) *Hayek on Hayek.* Chicago: University of Chicago Press.

Mirowski, P. (2002) *Machine Dreams.* Cambridge: Cambridge University Press.

Rizzello, S. (1999) *The Economics of the Mind.* Cheltenham, UK: Edward Elgar.

Ross, D. (2005) *Economic Theory and Cognitive Science.* Cambridge (MA): MIT Press.

Rowlands, M. (1999) *The Body in Mind.* Cambridge: Cambridge University Press.

Simon, H. A. (1947) *Administrative Behaviour.* New York: Macmillan.

Simon, H. A. (1979) *Models of Thought.* Vol. I. New Haven: Yale University Press.

Simon, H. A. (1983) *Reason in Human Affairs.* Chicago: Stanford University Press.

Simon, H. A. (1991) *Models of my Life.* Cambridge (MA): MIT Press.

Simon, H. A. (1996) *The Sciences of the Artificial.* Cambridge (MA): MIT Press.

Smolensky, P., G. Legendre, and Y. Myata. (1992) Principles for an Integrated Connectionist/Symbolic Theory of Higher Cognition. Hillsdale (NJ): Erlbaum.

de Sousa, F. (2005) Hayek and individualism: Some question marks. *History of Economic Ideas* 13 (2): 111–127.

de Vries, R. P. (1994) The place of Hayek's theory of mind and perception in the history of philosophy and psychology. In: J. Birner, and R. van Zijp (eds.) *Hayek, Co-ordination and Evolution.* London: Routledge, 311–322.

Weimer, W. B. (1982) Hayek's approach to the problems of complex phenomena: An introduction to the theoretical psychology of *The Sensory Order*. In: W. B. Weimer, and D. S. Palermo (eds.) *Cognition and the Symbolic Process.* Vol. 2. Hillsdale (NJ): Lawrence Erlbaum, 241–285.

THE RIGHT HEMISPHERE OF COGNITIVE SCIENCE

Bálint Forgács

Introduction[1]

The aim of the present study is to establish a theoretical connection between the brain (or more precisely the scientific concepts describing it) and the everyday expressions referring to the mental world. These expressions often circulate around dichotomies common in Western philosophy and thinking, like emotional–rational, mind–heart, or body–soul, the connotations of which are deeply embedded in language, although are often hard to notice. Still, they profoundly influence the perception, understanding, and interpretation of mental states. The main question is the following: Could the structure of such concepts originate from human cognition, and therefore from the architecture of the nervous system?

Independently from the philosophical question, whether the concepts addressing mental phenomena are somewhere "out" in the world – as in reductionism, e.g., Ryle (1949) –, or produced somehow "in" the mind – according to Berkeley's solipsism (Pléh 2000) –, there is a possibility that these dichotomies are a "byproduct" of our mental system. For example, the left and right hemispheres employ different sets of processes, like logical thinking as opposed to creativity (Hámori 2005) to address the very different task demands of the environment. These could be viewed as two fundamentally different perspectives on the world, both of which are well known to all of us, while only one of them is convenient to most individuals.

However, most concepts describing the three spatial dimensions of the nervous system seem to be rooted in the above philosophical opposition: emotion and reason (for right and left hemispheres), cognition and motivation (for cortex and limbic system), and action and perception (for anterior and posterior regions). On the one hand, this could be a confusing factor when theorizing neuroscience and during the operationalization of experiments; on the other hand, the recombination and a metaphorical interpretation of these labels might enable a new level of analysis. For example, the description of the anterior and posterior regions as being responsible for creating the balance of consciousness between the motor and sensory areas (involuntary actions and hallucinatory experiences in extreme cases –, see Fischer 1986), could be combined with the emotional–rational aspects of the two hemispheres. This way, brain researchers could pose questions from an unusual theoretical perspective.

Concepts describing the mental world are brought into scientific discussion from everyday thinking, and were linked to the brain through experimental observations, described also by everyday concepts. Therefore, scientists' explanations might reflect their preferred mode of

[1] Acknowledgements: I would like to express my gratitude for the invaluable guidance and help to Professors Csaba Pléh and György Bárdos.

interpretation, depending on their different perceptual, cognitive, and neural dispositions. For example, a personally convenient way of understanding might be reflected in taking representations either as visual images – or even as perceptual symbols (Barsalou 1999) –, or as language-like concepts (Fodor 1975), hence putting very different kinds in the focus of explanations. In a clinical context, the same process could motivate the idea that the core of human functioning is either emotional (as in psychoanalysis and humanistic psychology), or rational (as in behavior and cognitive therapy).

This problem is especially intriguing when the "mind" turns towards the "mind" itself – the mind, which possesses capabilities that are often very hard, or even impossible to describe scientifically – like fine art, creativity, or induction. Although there are known methods to operationalize creativity (Zétényi 2009) for example, and the right hemisphere is also known to be more creative (Hámori 2005), the question still lurks here: Are these concepts problematic neurologically or philosophically? Be this as it may, the latter is enough to have trouble with the former. Are these "hazy" concepts the result of human thinking: logical and sequential, as opposed to intuitive and unexpected? Are they the result of the neural architecture of the human mind and brain?

Metaphors and the brain

The first step of the study is to take a broad look on the linguistic structure of psychological concepts, and on the kinds of relations binding them together. The next is to try to assess how they relate to the brain. Mapping the connections of these different theoretical levels follows: The cognitive metaphor theory (Lakoff and Johnson 1980a) provides a plausible framework in the search towards a link between mental concepts and phenomenal experiences, which later opens a way to trace the neural systems producing them.

Metaphors and concepts

The cognitive metaphor theory of Lakoff and Johnson (1980a) proposes that metaphors are not ornaments of language, but the very building blocks of conceptual thinking. We understand abstract concepts by systematically mapping concrete concepts onto them. The easily comprehendible *source domain* (e.g., journey) is mapped on the abstract *target domain* (e. g. life). This works on a conceptual level (e.g., life is a journey), and can only be found in metaphorical expressions like "we had a bumpy year".

According to Lakoff and Johnson (1980b) only those concepts are not metaphorical which are derived directly from our experiences – concepts of orientation (up–down, in–out), ontological concepts (e.g., materials) and structured experiences (like eating, moving). The seemingly distant domains of metaphors are connected in their specific *experiential gestalts*: "a multidimensional structured whole arising naturally within experience" (Lakoff and Johnson 1980b: 202.). The motivation of metaphors is the basis of the mappings (e.g., the expression "he is a hothead" is motivated by "heat" and "anger" appearing in the same situation). Motivation can be closeness in time or space for example, although cultural aspects also play an important role. Hence it is impossible to foretell the metaphors of a certain language,

although one might tell which mappings are unlikely. These are the ones that are really counterintuitive to our very human experiences, like "anger" being "cold" (Kövecses 2002, 2005).

Grady (1997) divides metaphors into two groups: complex metaphors and primary metaphors. Complex metaphors are built up from primary metaphors that are closely related to our experiences (in the expression "warm smile", physical warmth and happiness are joint). In the case of primary metaphors sensorimotor and non-sensorimotor experiences get connected in a systematic way. Based on this idea, Lakoff and Johnson (1999) created the integrated theory of primary metaphors, according to which perhaps these mappings do not simply recall similar experiences, but activate the very same neural circuits. The research of Rohrer (2005) provided fMRI and EEG data that seems to verify this: while subjects read metaphorical sentences involving the hands (e.g., "hard to grab this idea"), many areas responsible for the motor control of the hands were activated. These results suggest that (primary) metaphors with experiential basis can be traced back to certain neural areas, and that we understand a great variety of knowledge domains by the activation and recombination of a relatively few neurocognitive sources.

Neurology's conceptual background

Words employed by researchers to describe the brain also possess an experiential background (many although indirectly). A number of them are not simply mental concepts – some refer to experiences directly like the labels of sensorimotor areas, and others to quite abstract concepts like "decision making". Nevertheless, all have been grounded to neural areas via experiments. So these words refer to some kind of experiential phenomena that was possible to link to a specific experimental situation. Still, these words carry connotations and broader meanings: Among the words having a psychological aspect, they have a location in the theoretical space of associations, connotations, even if at first sight it is not clear where.

Importantly, there is narrow set of words of psychological functioning that have their "location" in the human brain. Several possibilities follow from this. First, it is possible that the description of the brain somehow follows the structure of the words referring to the mental (this would be solipsistic stance); second, it is possible that there is simply no real relationship between the concepts of the mental phenomena and structure of the words labeling the brain (reductionist stance); thirdly, the correspondence is somewhat hazy, suggesting this question is not fruitful.

Mental metaphors

What kinds of metaphors hide behind psychological concepts? First, there is a kind of phenomenal orientation (the spatial would be extended by lighting and temperature), referring to the experiential environment of mental concepts. Here are some examples:

PRECISE THOUGHTS ARE BRIGHT, IMPRECISE THOUGHTS ARE DARK

"Bright mind!" "Clear thought." "I was enlightened by the speech."

PRECISE THOUGHTS ARE COLD, IMPRECISE THOUGHTS ARE HOT

"Cold calculation was the plan." "He has been a hothead with that decision."

POSITIVE EMOTIONS ARE BRIGHT, NEGATIVE EMOTIONS ARE DARK

"We had a brilliant time in the evening." "Dark intentions seized him."

POSITIVE EMOTIONS ARE HOT, NEGATIVE EMOTIONS ARE COLD

"His revenge was cold as ice." "She had warm feelings towards him."

Such basic concepts, based on mappings of primary metaphors, could combine, as subtle building blocks, into the metaphorically complex structure of abstract concepts. For example, irrationality is often linked to emotions, while thinking often seems rational, or mathematical proofs are thought to be objective, while subjective ideas often express attitudes or feelings. These more abstract concepts provide a kind of cognitive orientation in the mental space: even if their connotations are hiding in the background of associations, the connections between them talks about a complex framework. Since the connections are rooted in perceptual sensations, even the abstract domains might be linked to certain universal phenomenal experiences. Hence complex mental concepts like empathy or intuition of a highly abstract conceptual level might involve complex connotations derived from the lower levels of phenomenal orientations.

According to the previous analysis, psychological concepts might be a part of a mental space that refers to experiential grounds (both phenomenal and cognitive); nevertheless, they are the tools of scientists to describe the human brain in an objective manner. The aim of this study is to take a look at the metaphorical space of mental concepts with respect to neurology. Are there really connections similar to the previously described, among "phenomenal", "cognitive", and "conceptual" aspects of words referring to mental life? Do these correspond to actual neural regions of the brain?

Hypotheses

1) Psychological concepts can be arranged in a mental space representing the three spatial dimensions of the nervous system: the left and right hemispheres, the cortical and limbic systems, and the anterior and posterior regions.
2) Psychology major university students arrange these words differently compared to non-psychology humanities major students, thanks to their thorough elaboration.

Method

As a first step, a pilot study was designed to gather self-reported data on this matter. A questionnaire was created in which subjects had to assess psychological concepts.

Subjects

All together 83 people answered the test, 48 psychology majors, and 35 non-psychology, humanities majors, all graduate (MA) level university students.

The test

The questionnaire contained 105 words, each of which had to be assessed according to the dichotomies of everyday expressions referring to the three neural axes of the nervous system: thinking–emotion for the left and right hemispheres, consciousness–instinct for the cortex and the limbic system, and action–perception for the anterior and posterior regions. In each case, subjects had to decide whether the specific word fits the one or the other – there was no way to choose "neither" or "both". The 105 test words were a collection of the following:

1) Naïve psychological expressions *(heart, mind)*.
2) Scientific expressions of psychology *(cognition, reflex)*.
3) Expressions of sensorimotor orientation *(warm–cold, inner–outer*).
4) A kind of cognitive orientation *(subjective–objective, personal–social)*.

The latter was supposed to entail similarly coarse and general expressions as the sensorimotor orientation, although within a more conceptual domain.

Results

The data has been analyzed with the SPSS 17 software. A series of Pearson's chi-square was used to compare the two groups. Where the two groups did not differ, a second chi-square test (with 50% expected frequencies) was calculated on the whole sample, but in case the first test showed a significant difference, the second test was run on the two subgroups separately. Since all variables were analyzed two times, the level of significance was reduced to $p < 0.025$ according to the Bonferroni-correction.

The majority of the words were categorized the same way by the two groups: there were only 11 cases (out of the 315 decision), where only one of the groups could decide on the

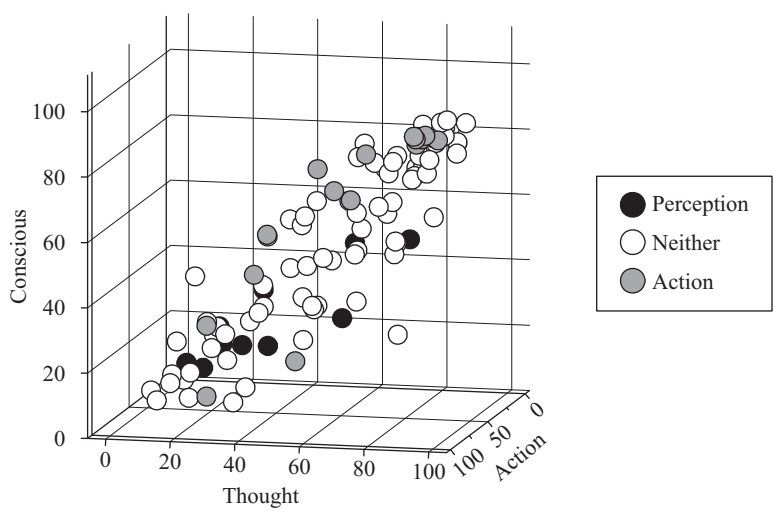


Figure 1. The 105 words of the research, in a three dimensional space, taking their average values on the three axes as coordinates. Each dot represents a word

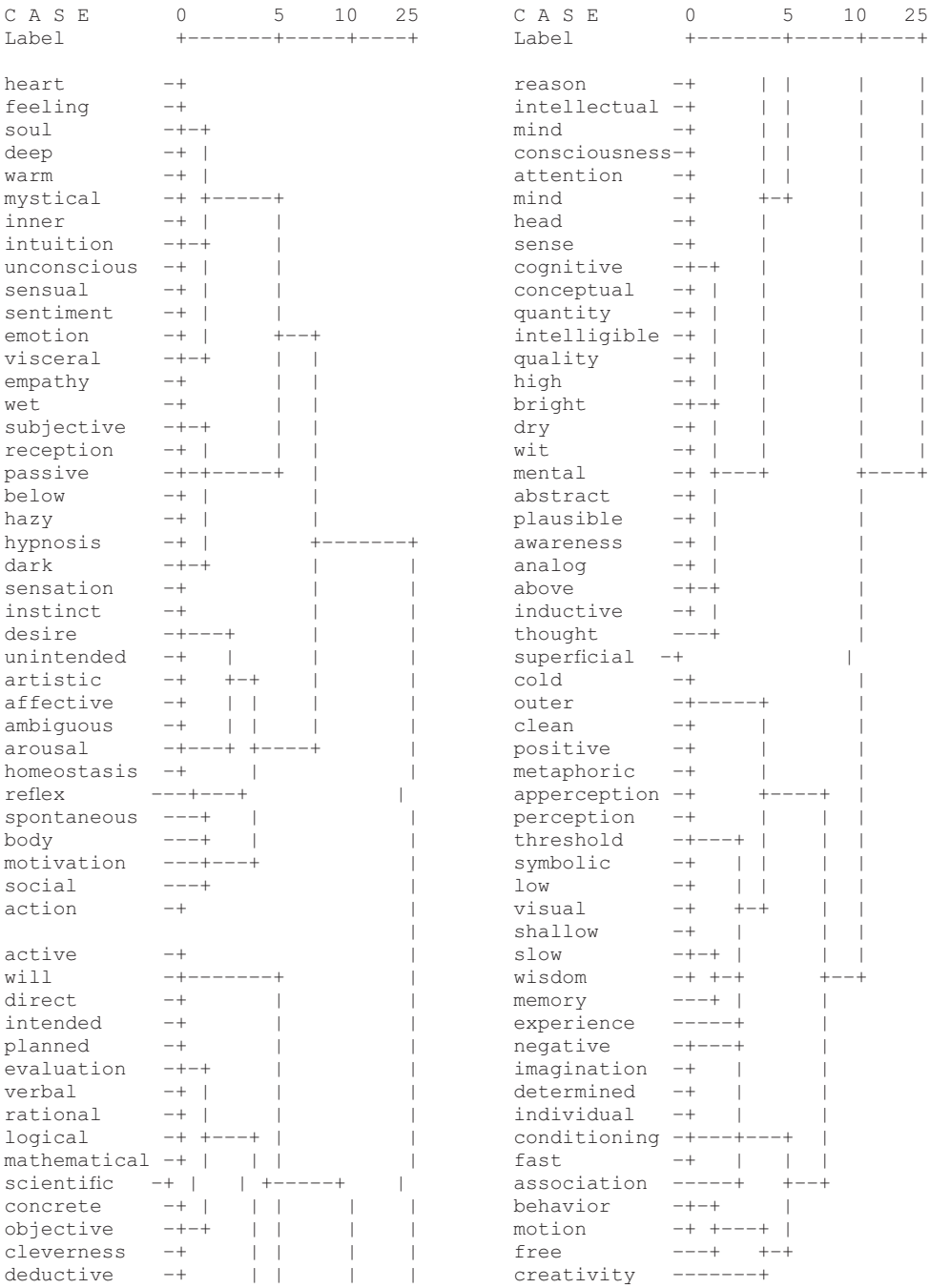


Figure 2. Results of a Hierarchical Cluster Analysis: Dendogram using average linkage

word, and only two words *(bright* and *quality)* were located on the opposite side of the ACTION–PERCEPTION axis. This approximately 4% difference was so small that these cases were considered as unsuccessful decisions, and were neglected. All words got either THOUGHT–CONSCIOUS or EMOTION–INSTINCT classifications except for one: the word *association* was categorized as INSTINCT and THOUGHT. Some words unexpectedly were categorized as EMOTION and/or INSTINCT, and ACTION at the same time (most were categorized as PERCEPTION once they got at least one of the former labels). These were: *reflex, social, creativity,* and *mystical.*

For the next step, the data was aggregated, thus the average answer for each word of the list was calculated. On this restructured data set, a factor analysis was run that revealed that the three dimensions actually fit onto one axis. One single component emerged, which is responsible for the 75% of all the variance in the sample. This can be seen on Figure 1, showing the words in a three dimensional space, according to their average values on the three hypothetical axes, taking the averages as coordinates.

Finally a cluster analysis was performed on the calculated averages of the three axes. The results are indicated on Figure 2.

Discussion

In general, the results of the present study are not strong evidence of any kind; nevertheless there are some interesting findings. First, the fact that psychology major university students did not differ significantly from non-psychology major university students of humanities indicates that the concepts used in the study (even the scientific ones) are deeply embedded in everyday thinking. This philosophical and naive psychological background certainly has some kind of influence on the conceptualization of research, and on the interpretation of results. Words expressing psychological phenomena bring their connotational net with them, and these might shape the understanding of mental life, and the human brain, since researchers, most of the time, have to choose from concepts with a history that cannot be neutral.

Another interesting finding was that the dichotomies corresponding to the dimensions of the brain do not differentiate sufficiently among the psychological expressions examined in the study – according to the factor analysis all of them actually fit onto one axis. This could be important for brain researchers, since it sheds light on the conceptual ambiguity of words used to describe very different levels of processes in the nervous system. Perhaps the philosophical mind–body problem appears here: even if it was often hard for individuals to make a decision on one word or another, the fact that finally the words got arranged according to one dichotomy suggests that the Cartesian dualism is deeply embedded in everyday and scientific thinking. This could be true even for scientists or philosophers, who actually deny being Cartesian. Another problematic aspect is the categorization this creates: For example, words that most people would consider "emotional" factually should also be linked to the "mind" instead of the "body". In general, this could be a confusing paradox of language to use in research, for all natural sciences exploring the psyche.

The cluster analysis revealed an intriguing structure among the concepts: They grouped together in accordance with the predictions of cognitive metaphor theory. For example, in the emotion cluster, words of perceptual orientation *(deep, warm, inner, dark),* and cognitive orientation *(subjective, spontaneous, active)* were located near to naïve *(heart, soul, body)*

and scientific psychological expressions *(empathy, unconscious* and *hypnosis* – as a Freudian slip). All the four levels were corresponding similarly on the other large cluster, the *mental* side too, as can be observed on Figure 2.

Of course, it is possible that the test was not constructed well enough, and the reason for only one emerging axis was that it did not differentiate enough (or did not produce enough differentiation), and so the gathered data is actually a research artifact. The reason why I chose such categorizing concepts was that all of them were commonplace in everyday language, but their meaning is taken for granted even by researchers. In this way, these categorizing concepts could bridge the gap between scientific and everyday discourse of mental phenomena.

Scientific metaphors in psychology

Based on this arrangement of the psychological concepts, and expanding the analysis to the mental world's conceptual space, it is possible to interpret the words describing the nervous system in a metaphorical map. The key concepts of various theories, or approaches in psychology, some of which are located in the brain in some way or another, might talk about the neural dimension central in the understanding of the specific theory. For example, the key concepts of psychoanalysis (e.g., *Libido, unconscious, instincts)* work as metaphors mapped onto a great variety of human functions. At the same time, these concepts can be read as metaphorically representing limbic level functions in the phenomenal mental space. In short: psychoanalysis is projecting limbic level functions on the whole brain. This analysis can be broadened to different schools, or simply to various theories in psychology. Such an arrangement of the various approaches of psychology, within the nervous system's phenomenal dimensions, according to their key metaphors can be seen on Figure 3.

Thus, researchers and their approaches could be identified according to their way of reasoning: the kind of neural processes they prefer to utilize during research, and the place where

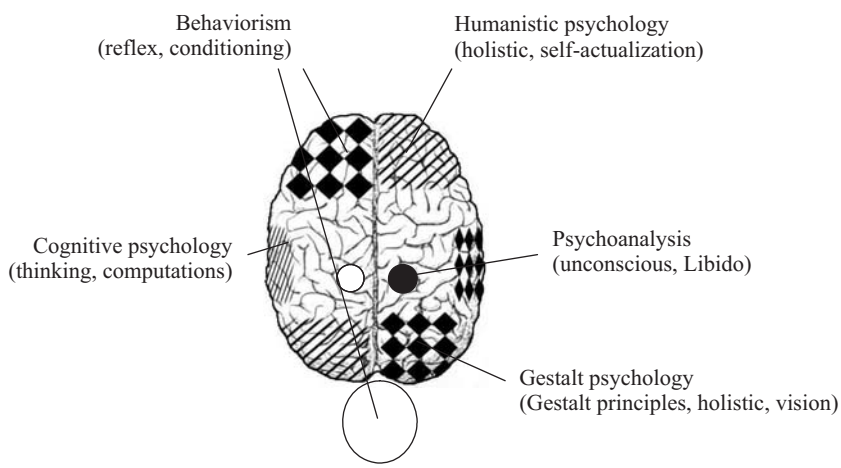


Figure 3. The different approaches of psychology, arranged via their key metaphors, according to the phenomenal dimensions of the brain

their scientific reduction collapses. This is a bit like a game with words, or a thought experiment, but could still be insightful. Continuing the above line of thought: The Gestalt principles refer to visually comprehensive, holistic ideas, which could be linked metaphorically to the right hemisphere's posterior regions.

The search for new paths in cognitive science

Taking a look at cognitive science from this perspective, an interesting picture emerges. Key ideas of early cognitive science like generative grammar or Turing machines talk about a left hemispherical, logical, sequential, and mathematical approach – while allowing mental functions to be assumed within natural science. Chomsky's (1957) ground breaking ideas approached from the linguistic domain, and syntax was (metaphorically) mapped onto the brain as computations (Fodor 1975). The brain was considered to be a special computer, where even emotions are "computed". Nevertheless, such "soft functions" of the mind, like empathy or creativity, proved to be very elusive, not just because it is very hard (or even impossible) to write an explicit protocol for them, but perhaps the main reason being that processes of one hemisphere cannot be mapped on the whole brain completely.

The interesting shift within cognitive science was that the need to take a look on the "missing" dimensions appeared as new currents within the established domain. As the first era lived up its theoretical resources, it was not a new paradigm that arose, but currents within the discipline that tried to bring in the "other side": connectionism (and pragmatism in general) offered models that do not work on rules, or computations in the classical sense, but the structure itself processes the information (Rummelhart and McClelland 1986). The architecture (the structure, the body?), and the procedure are considered prior to knowledge, metaphorically speaking, a shift from the declarative to the procedural. A similar thing happened (and was declared openly) when implicit processes became an independent field of research. Within cognitive linguistics, currents surfaced in the 1980s which tackled right hemispherical language functions as pragmatics (the Relevance theory of Sperber and Wilson 1985) and metaphor comprehension (Lakoff and Johnson 1980a). Figure 4 gives a brief summary of the idea. In general, the long-reigning research paradigm of the mental and psychological phenomena, cognitive science seems to be open enough to integrate approaches that do not

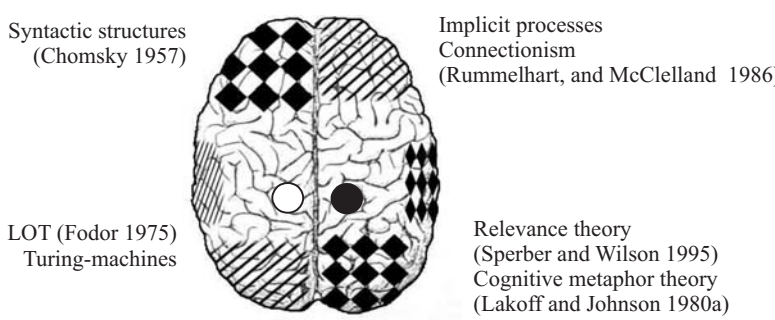


Figure 4. Approaches of Cognitive Psychology, arranged metaphorically in the brain

correspond closely to the original ideas – although the cognitive philosophy of the mind's doyen probably would not agree (Fodor 2008).

Epistemology, and the Brain for it

The main message of the work presented here is that there might be no single paradigm or approach to the understanding of the mental world, which could provide a fully consistent model. My argument is that any specific disposition or personal preference in understanding, interpreting and solving scientific problems concerning the brain might give scientists a hard time to produce a complete account of the very broad and diverse phenomena produced by the human brain itself. Of course, the tools of natural science are far too valuable to be abandoned, but certain aspects of the human mind could be almost impossible to describe logically or rationally. The brain might not be a computer in the sense we know computers. Two examples of these verbally hard-to-tackle processes are two mostly right hemispheric functions: sports and music. In both cases, when we are learning them, the teacher gives instructions that are dumb and vague if interpreted literally. Still, students do not just understand, but carry them out, sometimes even better than the teacher. There is less reason and more feeling to the mastering of such skills.

Even though this pilot study has also been fighting with several methodological problems, and is far from being perfect, I hope I could shed some light on questions accompanying psychological concepts and brain research. Perhaps the distance between the empirical and the theoretical aspects of this work can be bridged with time, and the gap could point towards the questions of research and epistemology, and the whole problem might bring more theoretical than empirical answers.

References

Barsalou, L. W. (1999) Perceptual symbol systems. *Behavioral and Brain Sciences* 22, 577–660.

Chomsky, N. (1957) *Syntactic Structures.* Berlin: Mouton de Gruyter.

Fischer, R. (1986). Toward a neuroscience of self-experience and states of self-awareness and interpreting interpretations. In: B. B. Wolman, and M. Ullman (eds.) *Handbook of States of Consciousness.* New York: Van Nostrand Reinhold C.

Fodor, J. A. (1975) *The Language of Thought.* Cambridge (MA): Harvard University Press.

Fodor, J. A. (2008) *LOT 2: The Language of Thought Revisited.* Oxford: Oxford University Press.

Grady, J. E. (1997) *Foundations of Meaning: Primary Metaphors and Primary Scenes.* Ph.D. Dissertation. University of California, Berkeley.

Hámori, J. (2005) *Az emberi agy aszimmetriái.* [The Asymmetries of the Human Brain.] Budapest: Dialóg Campus Kiadó.

Kövecses, Z. (2002) *Metaphor. A Practical Introduction.* New York/Oxford: Oxford University Press.

Kövecses, Z. (2005) *A metafora.* [Metaphor.] Budapest: Typotex.

Lakoff, G., and Johnson, M. (1980a) *Metaphors We Live By.* Chicago: University of Chicago Press.

Lakoff, G., and Johnson, M. (1980b) The metaphorical structure of the human conceptual system. *Cognitive Science* 4, 195–208.

Lakoff, G., and Johnson, M. (1999) Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought. New York: Basic Books.

Pléh, Cs. (2000) *A lélektan története.* [The History of Psychology.] Budapest: Osiris Kiadó.

Rohrer, T. (2005) Image schemata in the brain. In: B. Hampe, and J. Grady (eds.) *From Perception to Meaning: Image Schemas in Cognitive Linguistics.* Berlin: Mouton de Gruyter, 165–196.

Rummelhart, D. E., and McClelland, J. E. (1986) *Parallel Distributed Processing: Foundations.* Cambridge (MA): MIT Press.

Ryle, G. (1949) *The Concept of Mind.* London: Hutchinson and Co.

Sperber, D., and Wilson, D. (1995) *Relevance: Communication and Cognition.* 2nd ed. Oxford: Blackwell.

Zétényi, T. (2009) Lehet-e mérni az alkotókészséget? A kreativitás pszichometriája. [Can Creation Be Measured? The Psychometry of Creativity.] Paper Presented at the *Conference of Creativity and Talent.* Hungarian Academy of Sciences, 11 November 2009.

UNDERSTANDING THE RATIONAL MIND: THE PHILOSOPHY OF MIND AND COGNITIVE SCIENCE[1]

Olga Markič

Introduction

Bechtel, Abrahamsen and Graham proposed the following characterization of cognitive science: "Cognitive science is the multidisciplinary scientific study of cognition and its role in intelligent agency. It examines what cognition is, what it does, and how it works" (Bechtel, Abrahamsen, and Graham 1998: 3). If we take a look at the history of this relatively young science, we can see that scientists provide different answers to these three questions and are thus developing different research paradigms in cognitive science. In this paper, I will limit myself to a presentation of the interplay between scientific proposals about modeling and explaining rational agency, and questions and suggestions from the philosophy of mind. I will thus try to give a broader philosophical perspective to the different approaches provided by the classical and contemporary cognitive science research.

Setting the problem: Descartes' legacy

The modern Western view of the world can be traced back to the "Scientific Revolution" of the 17th century. At that time, the Aristotelian method of explanation in terms of final ends and "natures" was replaced by a mechanical, or mechanistic method of explanation in terms of regular, deterministic behavior of matter and motion. Galileo, Bacon, Descartes, and Newton introduced observation, experiment and the precise mathematical measurement as the main methods in the study of nature. This new "mechanical view of the world" was famously characterized by Galileo as follows: "This grand book of the universe, which […] cannot be understood unless one first comes to comprehend the language and to read the alphabet in which it is composed. It is written in the language of mathematics, and its characters are triangles, circles, and other geometric figures, without which it is humanly impossible to understand a single word of it" (Galileo, in Crane 1995: 3).

According to the mechanical view of the world, things behave as they do because they are caused to move in a certain way in accordance with the natural laws. This kind of explanation was used for inanimate nature, but many scientists wanted to explain living organisms with the help of the same method. Most famously, René Descartes thought that non-human animals are machines that could be explained from a purely mechanical perspective. We must bear in mind that in those times mechanical systems were taken to be systems which are de-

[1] I wish to thank Sebastjan Vörös for his helpful suggestions.

terministic and interact only on contact (e.g., a watch, as the most prominent metaphor). Later developments in science refuted the mechanical view of the world understood in this specific sense, but they haven't undermined the picture of the world where everything works according to the natural laws. I will follow the practice of Crane (1995) and Haugeland (1985), and use the term mechanical in this broader sense, synonymously with the nowadays more commonly used concept of "naturalistic".

The doctrine of vitalism, which presented an alternative to the naturalistic viewpoint concerning living things, has lost much of its previous appeal because of the empirical results in chemistry and biology. After the discovery of the DNA structure, it seems that even so complex a phenomenon as life can be, in principle, explained by mechanical processes. This leaves the mind as the most problematic phenomenon for the process of naturalization and leads us back to Descartes, who was willing to regard animals as mere machines that operate according to natural laws, but did not think the same about the human mind (soul). Descartes placed the mind *(res cogitans)* outside of the mechanical material world and adopted an interactionist dualist position concerning the mind–body relation.

It is often overlooked that Descartes in his *Traite del'homme (Treatise of Man)* provided a list of functions that can be explained without reference to the soul. This list contains not only functions belonging to the automatic nervous system, such as respiration and heartbeat, but also what we nowadays consider psychological functions, e.g., sense perceptions, memory and internal sensations like fear and hunger. Descartes thought that mechanistic explanation can be used to explain even such waking actions as walking or singing, when they occur without mental attention, but where mental attention is involved, we must posit a separate "rational soul" (Cottingham 1995: 146–147). Cottingham feels that "Descartes' mechanistic reductionism is starkly eliminative […] invoked only as a last resort, when the experimenter comes up against a phenomenon that baffles the explanatory powers of the scientist" (Cottingham 1995: 147).

In the *Discourse on method,* Descartes provides two arguments that deal with the issue of rational reasoning and support dualistic position from the scientific point of view. In his famous argument involving the human language user and his capacity to respond to an indefinite number of different situations, Descartes argues that machines are able to respond only with answers generated by the finite table that correlates inputs and outputs. It is thus "not conceivable that such a machine should produce different arrangements of words so as to give an appropriately meaningful answer to whatever is said in its presence, as even the dullest of man can do" (Descartes in Cottingham, Stoothoff, and Murdock 1991, Vol. 1: 140). Descartes' second argument is concerned with the universality of human mind which can be used in all kinds of situations, whereas physical systems "need some particular disposition for each particular action," and concludes that "it is morally impossible for a machine to have enough different organs to make it act in all the contingences of life in the way in which our reason makes it act" (Descartes in Cottingham, Stoothoff, and Murdock 1991, Vol. 1: 140).

These two arguments led Descartes to believe that reason could not feasibly be realized in a purely physical set of structures. Cottingham points out that the possibility of such a physical realization is not absolutely ruled out, and that Descartes as a good scientist was probably aware of this. Shanker (2004) in contrast stresses Descartes' repudiation of the doctrine of the 'Great Chain of Being' and his insisting "that there is a hiatus between animals and man that cannot be filed by any 'missing link'. The body may be a machine (which was itself a hereti-

cal view), but man, by his abilities to reason, to speak a language, to direct his actions, and to be conscious of his cognitions, is categorically *not* an animal" (Shanker 2004: 316). There is a bifurcation between mechanical, reflexive behavior and involuntary movements on the one side, and purposive behavior and voluntary movements on the other.

The view of Cartesian dualism stimulated many attempts to overcome the divide between the animals seen as mechanical automata on the one side, and rational human beings on the other. Shanker pointed out that the defense of the continuum picture could proceed in either of two directions: to show (1) that the behavior of animals is intelligent, or (2) that the behavior of man is mechanical. However much proponents of these two sides differ in what they consider as Descartes' legacy, they do take "Descartes as their spiritual leader" (Shanker 2004: 318).

The proponents of both approaches therefore accept the reality of mental phenomena (mental realism) but proceed from different starting points. The classical symbolic paradigm in cognitive science took the top-down direction, starting from folk-psychological notions and showing how mental states are physically realized. This marked the birth of cognitive science, "The Mind's New Science" (Gardner 1987). On the other side, there are attempts to provide an explanation by the detailed exploration of neural mechanisms (animal and human), a bottom-up approach directed to show how mental phenomena and rational behavior emerge as a product of evolution.

How is rationality mechanically possible? Cognition as computation

As mentioned, many philosophers, following Descartes, tried to find a place for the mind in nature by giving it a mechanical explanation.[2] For example, La Mettrie, an 18th-century materialist philosopher, provoked his contemporaries with the book the *Machine Man (L'Homme machine, 1747/1996).* For our purposes, the most interesting ideas were presented by Thomas Hobbes in *Leviathan* (1651), where he explicitly connected thinking and computing: "By RATIOCINATION, I mean *computation.* […] When a man reasoneth, he does nothing else but conceive a sum total, from addition of parcels; or conceive a remainder, from subtraction of one sum from another. […] These operations are not incident to numbers only, but to all manner of things that can be added together, and taken one out of another. For as arithmeticians teach to add and subtract in numbers; so the geometricians teach the same in lines, figures, […] angles, proportions, times, degrees of swiftness, force, power, and the like; the logicians teach the same in consequences of words; adding together two names to make an affirmation, and two affirmations to make a syllogism; and many syllogisms to make a demonstrations" (Hobbes, in Haugeland 1985: 24). Hobbes' idea that thinking is computing was not supported by the science of his time, and it had to wait until the advent of modern digital computers.

Descartes pointed out that one of the main problems of materialism is to explain how rational reasoning is possible. Haugeland (1985) called this problem "the paradox of mechanical reason", and characterized it as follows:

[2] This section is based on Markič (2004).

> Either the manipulator pays attention to what the symbols and rules *mean* or it doesn't. If it does pay attention to the meanings, then it can't be entirely mechanical – because meanings (whatever exactly they are) don't exert physical forces. On the other hand, if the manipulator does not pay attention to the meanings, then the manipulations can't be instances of reasoning – because what's reasonable or not depends crucially on what the symbols mean. In a word, if the process or the system is mechanical, then it can't reason; if it reasons, it can't be mechanical. (Haugeland 1985: 39)

The paradox of mechanical reason in a way repeats the dualist difficulty with the interaction. The way out proposed by the proponents of the representational theory of mind was a combination of a functionalist theory and analogy between the mind and computer. They based their approach on the proof theory in symbolic logic and took the computer as a model to show how to connect semantic properties of a symbol with its causal properties *via its syntax*.

The computational–representational theory of mind (CRTM) thus offers an explanation of how there could be non-arbitrary content relations among causally related thoughts, or, in Fodor's words: "How […] rationality [is] mechanically possible" (1987: 20).

CRTM and classical symbolic models are generally based on two claims (Fodor and Pylyshyn 1988: 12–13):

1) Representations have combinatorial syntax and semantics (the language of thought – LOT).
2) The principles by which representations are transformed are defined over structural properties of representations.

Sentences in LOT interact in a way that mirrors logical interactions of the contents. A token sentence of LOT is a concrete object that can have causes and effects and can therefore play a causal role. Mental symbols are in the individual's brain. "According to LOT, tokens of brain tissue in human beings – assemblies of neurons – constitute tokens of mental symbols having both semantic properties and syntactic properties" (Jacob 1997: 146). Fodor and Pylyshyn explain it in more detail as follows:

> The symbol structures in a classical model are assumed to correspond to real physical structures in the brain and the combinatorial structure of a representation is supposed to have a counterpart in structural relations among physical properties of the brain. For example, the relation 'part of', which holds between a relatively simple symbol and a more complex one, is assumed to correspond to some physical relation among brain states. […] The classical theory is committed not only to there being a system of physically instantiated symbols, but also to the claim that the physical properties onto which the structure of symbols is mapped are the very properties that cause the system to behave as it does (1988: 13–14).

Classical symbolic cognitive science provided a unified platform for interdisciplinary research based on the hypothesis that cognition is the processing of information. More precisely, information is encoded in the form of symbolic representations with rules operating upon them. Cognition works like computer. The investigation of cognitive processes focuses

primarily on activities where agents exhibit intelligent behavior (e.g., solving problems, reasoning, understanding sentences).

The relevance problem and the in principle/in practice distinction

Classical symbolic cognitive science was the first scientific approach that was on the verge of solving one of the great mysteries of the mind: *"How can its causal processes be semantically coherent?"* (Fodor 1987: 20) It seems that it has an *in principle* solution to the problem of mechanical rationality. Nevertheless, its attempt to explain rationality in a mechanical way is not without difficulties. The problem we will focus on is the relevance problem, or frame problem. It is hard to develop computational mechanisms that accurately model the human ability for everyday common-sense reasoning and decision-making. Human beings have an amazing ability to quickly see the relevant consequences of certain changes in a given situation. They understand what is going on, and are quickly able to draw the right conclusions. One of the problems with symbolic computational modeling is how to prevent fruitless time consuming and irrelevant inferences. The main questions the computational approach has to answer are thus the following:

1) How to deal with the changing world?
2) How to determine the relevant consequences of an event?

Scientists within the field of classical symbolic cognitive science try to solve the problem from different angles, exploring new non-monotonic logics, and new ways of knowledge representations. A strong line of criticism aimed at the classical symbolic cognitive science questions the primacy of declarative knowledge, and the neglect of the procedural knowledge ("knowledge how") that underlies our skills (Haselager 1997). It is interesting that Fodor, one of the main adherents of classical cognitive science, thinks that classical cognitive theory gives an *in principle* solution to the problem of mechanical rationality but is skeptical about the possibility of ever finding an *in practice* solution to the relevance problem.

The inability of finding an *in practice* solution to the relevance/frame problem shows that the top-down symbolic approach has serious limitations, and that, contrary to the initial optimism, it is unable to answer Descartes' concern about the possibility of designing a mechanical system that will behave reasonable "in all the contingences of life".

Difficulties with classical attempts to model reasoning and decision-making in everyday contexts contributed to the reassessment of answers related to the basic understanding of cognition. Cognitive scientists that were primarily engaged with research at the cognitive level (representations and algorithms), began to look to the brain for new inspiration. New connectionist and neural network models were developed and were quite successful in modeling previously neglected characteristics of human intelligent activity, such as graceful degradation, soft constraint satisfaction and generalization. But these new approaches to modeling are not unified in the same manner as it was the case with the symbolic paradigm. Some scientists argue that connectionist models are models at the implementational level and do not represent an alternative to the symbolic paradigm (Fodor and Pylyshyn 1988), while others take them as an alternative approach to modeling at the cognitive level and stress the

sub-symbolic character of connectionist representations (Smolensky 1988). There are those who take the dynamical system theory as a non-classical framework for cognitive science still employing representations (Horgan and Tienson 1996), and those who are more radical and think that the dynamical system theory explains cognition without representation (van Gelder 1995). They still describe cognition as information processing, but they understand it differently, discarding the computer/symbol manipulation analogy. They thus provide different answers to the question of *how cognition works.* Instead of the top-down approach, showing how intentionality can be naturalized via functional decomposition, these approaches (except for the implementationalist views) suggest that mental states emerge from networks of interconnected units (Bechtel and Abrahamsen 2002). Horgan and Tienson (1996) argue that their dynamical system theory approach provides better tools for solving the relevance problem because of its holistic features and abilities for generalization. I agree that connectionism is a better approach to solving the relevance problem, however, there is a serious possibility that difficulties will emerge, when "toy" models become bigger and more realistic, as it did so in symbolic modeling.

The lack of a unifying hypothesis provoked some sort of an "identity crisis" (Bechtel, Abrahamsen, and Graham 1998) among cognitive scientists at the end of the previous century. A plurality of approaches is still available, but it seems evident that cognitive neuroscience is now taking the leading role in providing explanations of how mental phenomena emerge from neural dynamics. Cognitive scientists are taking a broader perspective, and are investigating the mind by both the underlying brain mechanisms and relations to the body and environment (embodied and situated cognition). The researchers also no longer limit their investigations solely on processes that have been traditionally labeled as cognitive, but focus on emotions and will as well. In the next section, I will briefly present Damasio's suggestion about the role of emotions in decision-making and in solving the relevance problem.[3]

Broader perspectives: The role of emotions

The history of philosophy was dominated by the "negative view of emotions", and many philosophers (e.g., Plato, Descartes, and Kant) defended the view that in the process of rational decision-making, emotions were a hindrance to clear thinking. The prevailing theories of decision-making emphasized the rational aspect of the process. It was traditionally thought (Descartes' legacy) that the ability of rational decision-making was the main point of difference between human beings and other creatures. The role of emotions was in setting the goals and in motivation, whereas in reaching practical decisions, reason and emotion were in opposition. Recently, many philosophers, neuroscientists and psychologists have pointed out that emotions play an important role in decision-making on an unconscious level. For example, philosopher Ronald de Sousa claims that when dealing with the issue of decision-making, and especially with the relevance/frame problem, one can benefit significantly by accepting the hypothesis of emotions being active participants. He thinks that emotions make it possible that only a small percentage of all possible alternatives and facts become relevant in the process (de Sousa 2002).

[3] More in detail in Markič (2009).

Neurologist Antonio Damasio has come to similar conclusions, but from the perspective of neuroscience and psychology. He noticed that patients with damaged ventromedial prefrontal cortex had serious problems with decision-making in their everyday lives, but it was difficult to explain their irrational behavior because their abilities of rational reasoning seemed to remain intact. Damasio (1994 2003) started tackling this problem by investigating the mechanism which enables emotional processes to guide or direct decision-making. He named his approach the *somatic marker hypothesis.* With the help of his co-workers, he examined and tested patients with the damaged ventromedial prefrontal cortex (Bechara et al. 1994). He established that, when faced with situations which could lead to different types of action, these patients are unable to activate emotion-related memory that would help them pick the most advantageous alternative.

Damasio believes that an important part of the decision-making process consists of the comparison of potential alternatives with emotions and feelings from similar past situations. Furthermore, the process also involves the estimation of results brought about by these past events and potential rewards or punishments that might have been gained during such events. This procedure enables humans to simulate potential future outcomes based on their past experiences, and then opt for a move that will lead to the best possible solution. People tend to classify situations, experienced under the influence of social emotions and emotions of joy and sorrow, which are, in turn, triggered by rewards and punishments, in conceptual categories. These categories are, according to Damasio, formed on both mental and corresponding neural levels, and are later connected with brain apparatus responsible for triggering emotions. This enables appropriate emotions to come about quickly and automatically.

Emotional signals can be conscious and can make us re-live the corresponding feelings, or they can be hidden and automatic. In this case, emotional signals mark the possibilities and outcomes as positive or negative (a kind of alarm), and thus help us decide to take actions that are in accordance with our past experiences. Since such decisions are made relatively quickly and without conscious thinking, they are often called "intuitive". It is apparent that Damasio's theory is in agreement with de Susa's observation that emotions narrow the decision-making space. It gives us a possible explanation why reason alone, and therefore also classical symbolic cognitive science, fails to solve the relevance problem. It is this narrowing of the decision-making space that was lacking in Damasio's patients and in classical symbolic models.

Damasio's theory of emotions and feelings is very useful in explaining why normal people do not face the relevance problem, while those who lack emotional feelings have difficulties deciding in uncertain situations. Damasio has given a strong push to further investigations of the role of emotions in decision-making, and to the examination of networks of different brain areas that are involved in cognition and emotion. At the same time, we have to be aware that Damasio's theory is not devoid of certain deeper-level problems. I think the most pressing one is his distinction between feelings and emotions, where the first are mental and the latter not. As I see it, this is not merely a matter of methodological separation, but points to the old mind–body problem. Although Damasio tries to show that his solution is in accordance with Spinoza's double aspects monistic position, where mental and physical are parallel aspects of one substance, he often slips in the interactionist conceptualisation of physical and mental (Damasio 2003). If my observations are correct, Damasio's theory fails to provide a plausible explanation of how processes where we need mental attention are "just mechanical".

Conclusion

I have tried to show the dialectic between the philosophical/theoretical and the empirical. Scientists from different paradigms and disciplines of cognitive science are trying to answer Descartes' challenge and provide models and theories of mechanically explaining rational behavior. It seems that classical computational approach that is based on the computational representational theory of mind provides *in principle* solution, but fails to do it *in practice.* On the other hand, cognitive neuroscientists who investigate neural correlates of mental states are often not cautious enough in their interpretations and overlook the trap of the mind–body problem. Of course, one could say that the main problem lies in the way Descartes articulated the problem and in his acceptance of the reality of the mental, but that's already different story.

References

Bechara, A., Damasio, A. R., Damasio, H., and Anderson, S. W. (1994) Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition* 50, 7–15.

Bechtel, W., and Abrahamsen, A. (2002) *Connectionism and the Mind.* Oxford: Basil Blackwell.

Bechtel, W., Abrahamsen, A., Graham, G. (1998) The life of cognitive science. In: W. Bechtel, and G. Graham (eds.) *A Companion to Cognitive Science.* Malden, Oxford: Blackwell Publishers.

Cottingham, J. G. (1995) Cartesian dualism: Theology, metaphysics, and science. In: J. G. Cottingham (ed.) *The Cambridge Companion to Descartes.* Cambridge: Cambridge University Press.

Cottingham, J. G., Stoothoff, R., and Murdock, D. (eds.) (1991) *The Philosophical Writings of Descartes.* Vol. 2. Cambridge: Cambridge University Press.

Crane, T. (1995) *The Mechanical Mind.* London: Penguin Books.

Damasio, A. (1994) Descartes' Error: Emotion, Reason, and the Human Brain. New York: G. P. Putnam's Sons.

Damasio, A. (2003) Looking for Spinoza: Joy, Sorrow, and the Feeling Brain. London: William Heinemann.

Descartes, R. (1664) *Traite del'homme* [Treatise of Man.] Paris: Angot.

Fodor, J. A. (1987) *Psychosemantics.* Cambridge (MA): MIT Press.

Fodor, J. A., and Pylyshyn, Z. W. (1988) Connectionism and cognitive architecture: A critical analysis. *Cognition* 28, 3–71.

Gardner, H. (1987) *The Mind's New Science.* New York: Basic Books.

Haselager, W. F. G. (1997). Cognitive Science and Folk Psychology: The Right Frame of Mind. London: Sage Publication.

Haugeland, J. (1985) *Artificial Intelligence: The Very Idea.* Cambridge (MA): MIT Press.

Hobbes, T. (1651) *Leviathan.* London: Andrew Crooke.

Horgan, T., and Tienson, J. (1996) *Connectionism and the Philosophy of Psychology.* Cambridge (MA): MIT Press.

Jacob, P. (1997) *What Minds Can Do.* Cambridge: Cambridge University Press.

La Mettrie, J. (1996) *Machine Man and Other Writings.* In: A. Thomson (ed.). Cambridge: Cambridge University Press.

Markič, O. (2004) Mechanical rationality. In: A. Ule, M. Gams, G. Repovš (eds.) *Proceedings A of the 7th International Multiconference Information Society. IS 2004. October 9–15, 2004. Ljubljana, Slovenia.* (Informacijska družba.) Ljubljana: Institut Jožef Stefan, 137–139.

Markič, O. (2009) Rationality and emotions in decision making. *Interdisciplinary Description of Complex Systems* VII (2), 54–64.

Shankar, S. (2004) Descartes' Legacy: The mechanist/vitalist debates. In: S. Shankar (ed.) *Philosophy of Science, Logic and Mathematics in the Twentieth Century.* London, New York: Routledge.

Smolensky, P. (1988) On the proper treatment of connectionism. *Behavioral and Brain Sciences* 11, 1–74.

de Sousa, R. (2002) *Why Think? Evolution and the Rational Mind.* Oxford, New York: Oxford University Press.

van Gelder, T. (1995) What might cognition be, if not computation? *Journal of Philosophy* XCII (7), 345–381.

FROM COGNITIVE CARTESIANISM TO COGNITIVE ANTI-CARTESIANISM: A HYPOTHESIS ABOUT THE DEVELOPMENT OF COGNITIVE SCIENCE

Jean-Michel Roy

1. A plea for the history of cognitive science

Cognitive science, understood as the specific attempt to develop a scientific theory of cognitive faculties born in the mid-1950s and carried out until the present time, by and large still awaits its historians. Historical accounts of its emergence and unfolding remain indeed very rare – to the extent that we even have one.

This judgment might sound too severe. It nevertheless depends on what one means by a historical account, as well as on what one deems it legitimate to expect in this respect in the area of contemporary cognitive theorizing. If we think of a historical account of sufficient breadth, level of factual details, theoretical depth, and methodological rigor to compare with what several other sectors of the modern history of scientific knowledge now have to offer, this negative assessment can hardly be disputed.

In the first place, comprehensive coverage of the cognitive science enterprise of a satisfactory amplitude is virtually absent. The voluminous *History of Cognitive Science* of Margaret Boden (Boden 2006) is probably the only work to really qualify as such today. And it is symptomatic that it took twenty years to have a replacement for the ground-breaking *History of the Cognitive Revolution* of Howard Gardner (Gardner 1985) that had been inexorably reduced, through the course of years, to the status of an opening chapter of such a comprehensive history. Denouncing the quasi absence of comprehensive histories of appropriate magnitude in no way amounts to denying the existence and significance of very suggestive historical essays that confront the task of covering the whole development of cognitive science, such as the opening chapter of *A Companion to Cognitive Science* (Bechtel et al. 1998), or the introduction and conclusion of *Introduction aux Sciences Cognitives* (Andler 2004). However, not only are such historical essays themselves very few in number, they also are no more than blueprints of full-fledged histories.

In addition, be they simple blueprints or full-fledged historical attempts, existing accounts fall in the category of what is known as immediate history, rather than history *per se.* Indeed, they are the products of protagonists of the cognitive enterprise who, however impressive the erudition they can manifest, are not professional historians of science sufficiently detached from their object of inquiry, and sufficiently aware of the technicality of the modern history of scientific ideas. Caesar's account of the Gauls' wars, however interesting and valuable that it might be, remains the testimony of one of the warriors more than a piece of real history.

Finally, if authentic historical works undeniably exist in the area of cognitive science, they are in fact monodisciplinary ones, dealing with the sole development of linguistics (e.g., Newmeyer 1986, 1996), psychology (e.g., Lachman et al. 1979, or Baars 1986), artificial

intelligence (e.g., McCorduck 1979), neuroscience (e.g., Finger 1994, or Clarac et al. 2008), or cognitive neuroscience (Bennett and Hacker 2008). Although these analyses undeniably constitute crucial materials for the elaboration of a history of cognitive science, most of them nevertheless lack the integrative perspective that is so essential to the cognitive science movement, in order to be seen as authentic steps in the construction of a history of this movement. As a matter of fact, such historical investigations develop as independent sectors.

Surprisingly, the fact that the history of cognitive science remains such a *parent pauvre* of the modern history of science does not elicit much reprobation. The idea seems still pervasive that cognitive science is too young a discipline to offer a proper domain of inquiry to historians. Young it certainly is, when compared with most other disciplines, but not enough to conceal its true age. Now a 60-year-old cognitive science is by no way a newborn, even if its name still sounds unfamiliar to the general public. And its life has been so far a quite important and eventful one by any count, making it on the contrary a privileged object of interest for the historian. There is consequently no reason why modern history of science should continue to wait until it grows more before starting to consider it as a topic worth its most dedicated attention.

Furthermore, there are in fact many reasons pleading in favor of an urgent repair of this undue neglect, and in particular of the fact that cognitive science itself needs a real history of its development. A scientific enterprise is indeed like an expedition, and it cannot be appropriately conducted without a clear knowledge, at any moment of its course, of where it is supposed to be, where it is supposed to go, and how it is supposed to move from the first point to the second. And this knowledge will be all the more reliable if it is rooted in a no less clear memory of where the enterprise started, where it was initially heading for, and how it got to the place where it is at any such moment.

Three illustrations at least of the usefulness of a careful historical analysis can be mentioned in the specific case of the cognitive science enterprise.

Misconceptions of its past remain widespread, and even seem to expand as the cognitive revolution out of which it is commonly supposed to be born becomes more distant, favoring misapprehensions of its theoretical identity that generate false oppositions and facilitate shaky claims to novelty. For example, by considering early cognitivism as a kind of more or less implicit dualism, many partisans of the neuro-cognitive approach to cognitive explanation that emerged in the 1990s ignore that naturalism is probably the most central principle of the cognitive revolution. And confuse the fact that cognitivism might be a failed attempt at cognitive naturalism with the fact that it would be one only by courtesy, thereby weakening their own case through an obvious misapprehension of their target. Another familiar misconception is that cognitive science is a new *discipline,* while the very idea of elaborating a scientific theory of cognitive faculties runs deep in the whole development of western knowledge, and is at the heart of the philosophical tradition. And acknowledging this connection is essential to a proper assessment of its innovations. An additional example of the distorted view of the past of cognitive science favored by the lack of a strong field of investigation of its development is that, from a philosophical point of view, cognitive science is believed to belong intrinsically to the analytical current. While it does not according, for example, to the most well-known definition of analytical philosophy due to Michael Dummett, because it reinstates an architecture, as clearly exemplified by philosophers like J. Fodor, or J. Searle, where the philosophy of mind is theoretically prior to the theory of language (Roy 1999).

Knowledge of the development of cognitive science is not only needed to prevent errors about what it was and might still be, but also to dissipate the obscurity still surrounding the processes that led to the formation of its first figure, that of cognitivism, and subsequently drove it away from this cognitivist figure to its current, and very different, situation. If the main theoretical sources of the cognitive revolution are well identified (calculability theory, mathematical communication theory, neural networks theory, cybernetics, etc.) and studied (cf. e.g., Tête and Pélissier 1995, and Triclot 2008), the way how these different ideas merged into the project of an integrated theory of cognitive faculties remains indeed in many respects to be accounted for (cf. however Dupuy 2000), as well as the moment when this integrated theory can be legitimately declared to be born. The date of 11 September 1956 popularized by the psychologist George Miller (Miller 1979), and corresponding to the second day of an MIT conference that gathered many of its decisive contributors and when one could for the first time perceive the strong commonalities between their respective projects, is highly symbolic. Moreover, the very complex process of transformation that accompanied the growth of the newborn discipline and resulted in its fragmented and heterogeneous landscape face is even more in need of a precise reconstruction, in spite of the heroic effort represented by M. Boden's 2006 book, and the wealth of details that it offers in this respect; and also in spite of a few precious essays that attempted to analyze from a non-historical perspective the theoretical changes that took place since the early days of the cognitive revolution, such as Andy Clark's 1997 *Being There* in particular (Clark 1997).

Finally, this reconstruction task seems all the more important in view of the state of undeniable confusion that cognitive science has now reached. It is indeed far from clear that anybody today has a decent understanding of the extremely intricate architecture that the field has developed since the 1990s. An intricate architecture due not only to an explosive flourishing of disciplines and sub-disciplines, and to a correlative uncertain redrawing of boundaries between research areas, but, more importantly, to a multiplication of general research programs (dynamical cognitive science, enactive cognitive science, embodied cognitive science, embedded cognitive science, situated cognitive science, phenomenological cognitive science, neuro-cognitive cognitive science, etc.) which has transformed cognitive science, from a nearly monolithic enterprise committed to the principles of cognitivism, into a field of intense competition between innumerable theoretical rivals. To the point that it is now exploring so many different directions that it has become urgent to clarify what these directions really are and how much they really differ one from the other. In other words, to try to get a decently accurate picture of where cognitive science research is really going, and on the basis of such clarification, to determine where it *should* really go. Undeniably, a crucial instrument of this clarification of its present situation and of its ideal future is the possession of an accurate and detailed understanding of the path so far followed.

2. The principles of a would-be account

Speculation hardly seems the proper way to reach such an understanding, unless it is conceived as the formulation of a hypothesis, suggested by a number of observations and theoretical considerations, to be used as a heuristic guide in the rigorous exploration of historical facts and open both to further articulation and to empirical testing. And it is in

such a form of speculative history of cognitive science that I would like to briefly engage in the remaining pages. Offering a full-blown speculative hypothesis about the development of cognitive science would require presenting a comprehensive picture of the main stages of this development from the years of its formation to its most recent decisive transformations. Understandably, I will here limit my ambition to lay out and defend some of the most important principles upon which such a would-be account, to use the felicitous terminology of the editors of this volume, should in my opinion be based.

Hypothesis One: The need for a theory-based history of cognitive science

It seems clear, in the first place, that a solid historical account of cognitive science cannot ignore the theoretical issues dealing with the mode of development of scientific ideas, and should either capitalize on some of the answers proposed for them by the contemporary history of science (T. Kuhn, I. Lakatos, etc.), or argue for the inadequacy of these answers and elaborate more appropriate ones. Similarly, it should operate with a clear set of epistemological principles and categories in its analysis of historical facts, maintaining, however, an adequate level of generality to be able to capture the specificity and variety of epistemological standpoints within cognitive science itself, as well as the transformations that it undergoes in this respect. In short, a solid history of cognitive science must be based both on a theory of scientific knowledge and on a theory of its mode of evolution. What is at stake in the elaboration of these theoretical underpinnings is the meaning of the historical facts that it investigates.

Hypothesis Two: The uncertainty of the cognitivist revolution

The fact that this is so is best illustrated with the idea of a cognitive revolution. Introduced by H. Gardner, the idea has gained wide currency but seldom corresponds to any real concept, in spite of the various elaborations that the notion of scientific revolution has received in the modern theory of the mode of development of scientific ideas. And it is far from clear that it can survive its subsumption under one of the technical concepts produced by such elaborations. For example, it is at times referred to the Kuhnian notion of scientific revolution as paradigm shift. The shift being seen in this case as one from behaviorism to cognitivism. But the existence of a cognitivist paradigm is disputable. Indeed, Kuhn explicitly designates with this notion a singular achievement capable of rallying a whole scientific community. Was there any work playing such a role in cognitive science? The most one can find is a series of disciplinary paradigms. Chomsky's *Linguistic Structures* (1957) is the best-fitting case, but the situation of psychology is more delicate: Is such a paradigmatic role to be attributed for instance to G. Miller's *Plans and the Structure of Behavior* (1960), or to the later A. Newell and H. Simon's *Human Problem Solving* (1972)? For decisive that they were, their earlier 1956 contributions, "The Magic number seven" (Miller 1956), or "The Logic Theory Machine" (Newell and Simon 1972) seem to be too limited. At any rate, the very idea of a plurality of disciplinary paradigms does not fit the Kuhnian concept of scientific revolution, and the Kuhnian theory does not seem to offer the adequate resources to capture the element

of integration that lies at the core of the scientific transformation that affected the field of cognitive studies in the mid-1950s. Interestingly, in one of his presentations of the cognitivist conception of cognitive science, J. Fodor underlines its anonymity, antagonist with Kuhn's idea of scientific achievement:

> Generally speaking, theories have certain logical properties with headaches. For example, every theory, like every headache, tends to be somebody's theory or other. […] The theory I want to tell you about […] is an exception to this rule. It has, so far as I can tell, no definite provenance, having developed as a sort of epiphenomenon of work in a number of different fields; most notably linguistics, computer science, logic, psychology and philosophy. In consequence there isn't any place you can go for the official, authorized version. […] It is now widely agreed that, whatever exactly the theory is, the right thing to call it is: cognitive science. (Fodor 1987: 105)

For the same reason, as also emphasized by Fodor, cognitivism has many different versions, a fact compatible with Kuhn's theory, but associated with a paradigm crisis, not with a paradigm's constitution.

Questioning the relevance of the idea of a cognitive revolution is certainly not to deny the reality of the emergence of cognitivism. But it seems preferable to analyze this emergence primarily in terms of the constitution of a specific cognitive foundational hypothesis, understood as a set of answers to the most basic problems that a theory of cognitive faculties must address, such as the delineation of its domain, the right conception of scientificity, the choice of fundamental concepts and principles, etc. One particularly important aspect of this foundational hypothesis being the integrative stance it adopts in the definition of its domain.

In this perspective, the problem of the existence of a cognitive revolution becomes essentially that of determining how new that cognitivist foundational hypothesis is, and what kind and degree of novelty from a scientific hypothesis is required for its emergence to count as a revolutionary transformation. A reasonable assumption is that this novelty must be located at a very fundamental level of the theory. And in this respect, the cognitivist hypothesis qualifies as a revolutionary candidate, since it pretends to redefine the very foundations of the scientific knowledge of cognitive phenomena. However, it is less clear that the content of this redefinition is sufficiently innovative to be labeled revolutionary in any convincing sense of the term. As emphasized by some of its best and fiercest partisans, such as Fodor or Chomsky, the cognitivist hypothesis is indeed largely a revival of a 17th- and 18th-century way of conceiving cognitive faculties, as adamantly denounced by neo-Wittgensteinians. There is consequently no shortage of reasons for emphasizing its conservative side, and seeing its rebellious overthrow of the behaviorist principles of explanation more as a restoration than as a revolution. In the same article, Fodor writes, again revealingly: "I shall be emphasizing… respects in which cognitive science is a recidivist account of the mind. In my view, much of cognitive science is the rediscovery of doctrines that were familiar in the tradition of classical epistemology… a curiosity of the cognitive science movement is that it has significantly more in common with the psychological theorizing of the 17th, 18th and 19th centuries than it does with the psychological theorizing of the first half of the 20th century…" (cf. Fodor 1987: 105). One neglected aspect of this recidivist dimension is the fact that, from a philosophical point of view, cognitivism can be seen as the expression of a larger cognitive turn within the

analytical movement, one that consists in reasserting the theoretical priority of the philosophy of the mind over the philosophy of language,[1] as well as reuniting the philosophy of mind with the intentionalist perspective stemming from Brentano. This is not to say that cognitivism does not bring new theoretical ideas, or a new twist to old ideas, but simply to question that these innovations are of a sufficient degree of novelty to transform it into a revolutionary hypothesis.

Hypothesis Three: A process of foundational revision

This suspicion in no way condemns cognitive science to utter conservatism. On the contrary, it opens the possibility that the real cognitive revolution came at a later stage, or even just started to point on the horizon. This possibility is intrinsically related to another striking feature of the development of cognitive science, namely that it has so far been accompanied by a constant revision of its early foundations. Far from being a cumulative process of complexification and readjustment of the founding principles of cognitivism, it is undeniably marked by recurrent attempts to introduce new foundations that conflict with them in a varying proportion, giving birth to alternative views of what a scientific account of cognitive phenomena amounts to at its most basic level, such as the various ones previously mentioned – a characteristic also in contradiction with the Kuhnian idea of a scientific revolution, which requires that cognitivism opened a period of paradigm-based research. The earliest claim to such a foundational transformation came in the mid-1980s with the revival of connectionism and the switch to a different conception of the fundamental notion of computation, but most of them crystallized in the 1990s, even if some only flourished in the 2000s.

Hypothesis Four: A systematic challenging of cognitivism

In fact, this process of foundational revision can be reconstructed, even if it was not always conducted in such spirit, as a systematic questioning of the defining principles of cognitivism. Although it is well-established that it more or less directly targeted cognitivism, the systematicity and precision of the anti-cognitivist dimension of this process of revision has been obscured by an overly limited understanding of cognitivism, too often reduced to a combination of classical computationalism and representationalism. Cognitivism is, however, a much more comprehensive hypothesis, in which the implicit limitations put on cognitive explanation are as important as the explicit principles imposed on it. Its core version can be briefly defined as the conjunction of the following main ideas:

1) *Internalism:* Cognitive behavioral manifestations are to be explained by postulating causal processes inside the behaving cognitive system.
2) *Naturalism:* These internal processes are natural ones, in the sense of physico-biological ones.

[1] And for this reason considered by Dummett as betraying its true identity.

3) *Abstraction, and non-reductionism:* However, they should be studied at an abstract level, different from, but also ontologically dependent on, the implementation level dealing with their physico-biological dimension. This abstract level defines the specific domain of cognitive science.
4) *Heuristic independence:* Hypotheses of the abstract or cognitive level are not to be constrained by specific hypotheses of the implementation one.
5) *Mentalism:* The cognitive level partially corresponds to what common sense and the psychological tradition calls mental processes.
6) *Representationalism:* Cognitive processes involve states with representational capacities.
7) *Intentionalism:* Representational capacities are identical with intentional ones.
8) *Symbolism:* Being representational is also no different from being symbolic.
9) *Classical computationalism:* Cognitive processes are also Turing's computational ones.
10) *Exclusion of consciousness:* The study of conscious manifestations accompanying cognitive processes is banned.
11) *Cold cognition:* Similarly, mental processes of an emotional kind, sharply distinguished from cognitive ones, do not fall into the purview of cognitive science.
12) *Rejection of first person methods:* The behaviorist rejection of any form of first person method is maintained.
13) *Body inessentialism:* Body plays no essential role in cognitive processing.
14) *Action inessentialism:* Action is not recognized as a cognitive phenomenon which is more basic than others.
15) *Situation inessentialism:* The environment is reduced to a source of distant stimuli, and a recipient of behavioral outputs, of internal cognitive processes.
16) *Methodological solipsism:* The environment does not exert any essential constraint on cognitive processes.

In my opinion, it can be shown that virtually every major transformation that marked the development of cognitive science since cognitivism, not only is a foundational one, but puts directly into question one or more of these principles. The emergence of cognitive neuroscience is, for instance, a rejection of point 4, the "consciousness boom" a refusal of point 10, enactive cognitive science at least a disagreement with point 14, the phenomenological approach to cognitive science (according to its central version) a reversal of point 12, etc.

Hypothesis Five: A state of foundational crisis

A further contention is that none of the theoretical conflicts opened by these anti-cognitivist attacks, with the possible exception of the cognitivism versus connectionism issue, has yet found a reasonably consensual solution, thereby plunging cognitive science into an unprecedented state of foundational uncertainty since it emerged. As a matter of fact, in a striking contrast with the cognitivist orthodoxy that directed its first steps – with some of its staunch defenders nevertheless still vigorous (Fodor 2009) –, a whole array of apparently heterogeneous directions are now being explored within cognitive science. This heterogeneous re-orientation demonstrates unquestionably that a first form of theoretical unity has been abandoned, and that no substitute has yet been found, leaving thereby, below the surface

accumulation of empirical results, most of the foundational issues of a science of cognition once again unresolved.[2]

Hypothesis Six: The anti-Cartesian drive

Having not only rejected its prior cognitivist direction, but also found apparently no other real one, a legitimate question consequently arises about the future development of cognitive science: where is it going to?

A final element of the speculative hypothesis sketched out here is that a closer look at its past might provide some clues as to what the answer might be.

It is arguable, indeed, that the systematic pulling apart of cognitivism obeys, to a large extent at least, a deeper logic of uprooting a general conception of the nature of cognitive systems considered as typically Cartesian because of finding its paradigmatic expression in Descartes. And consequently, that most if not all counterproposals to cognitivism are efforts to free cognitive science from these putative Cartesian chains, revealing the search for a non-Cartesian conception of cognitive faculties as their minimal common denominator. In other words, the moving away from cognitivism is commanded by an anti-Cartesian drive that gives it its deepest theoretical unity, putting into light that it fundamentally consists in moving both away from cognitive Cartesianism and towards cognitive non-Cartesianism.

That it might be so comes as a surprise only to those with a superficial knowledge of cognitivism, who are aware of its rejection of substance dualism but who ignore its conservative side. A conservatism that amounts not only to a restoration of general views cutting across the philosophical spectrum of the classical age, but also to some specific Cartesian options such as nativism, epitomized by Chomsky's concept of Cartesian linguistics. However, the anti-Cartesian drive is less directed at explicit vindications of Cartesianism than at what it sees as unconsciously entrenched Cartesian prejudices in need of being overturned in a way that is reminiscent of more than one deconstructive strategy. It is less a question of going against cognitive Cartesianism than of going beyond cognitive Cartesianism.

The main argument supporting this hypothesis is the recurrence, throughout the development of cognitivism, of explicit anti-Cartesian claims and calls to go beyond Cartesianism. Their diversity of object, context, target, and degree of elaboration has tended to mask their quantitative importance as well as their commonality of perspective. But a few examples suffice to show that they should not be seen as peripheral and scattered reflections.

The most well-known ones are probably due to D. Dennett, who insisted on dissociating the re-integration of consciousness as an object of inquiry from its association with what he calls a "Cartesian theatre" (Dennett 1991), and to Antonio Damasio, who saw the "affective turn" as a victory of Spinozist ideas over deeply ingrained intellectualist tendencies inherited from Descartes (Damasio 1994). But it is also H. Putnam arguing in the 1990s that cognitive science remained prisoner to a "Cartesian cum materialist picture" of perception that wrongly prevents it from accepting the possibility of a direct apprehension of the world by a cognitive system (Putnam 1999); or J. Searle defending the view that the core issue of naturalism is a pseudo problem because it is based on an unnoticed and unacceptable "conceptual dualism"

[2] For further argumentation of this point, see Roy (2010).

of Cartesian origin (Searle 1992). Another illustration is the denunciation by M. R. Bennett and P. M. S. Hacker (2003, 2008) of the "Crypto-Cartesianism" of the neuro-cognitive alternative itself to cognitivism, resulting in their eyes from the reproduction of the "mereological fallacy" of Descartes, which consists in attributing to a sub-part of a cognitive system properties that only belong to it as a whole. Examples can be multiplied, but the most comprehensive denunciation that "orthodox cognitive science is Cartesian in character" is probably contained in M. Wheeler's *Reconstructing the cognitive world* (Wheeler 2005), which also contends that overcoming these Cartesian limitations is what the situated and embodied approach is after.

3. Cognitive Cartesianism: How far beyond?

The correctness of the view laid out in the previous pages is to be determined by a detailed historical investigation of the development of cognitive science. Beyond empirical testing, it also stands in need of further articulation, especially regarding the concepts of cognitive Cartesianism and anti-Cartesianism which it involves. One obvious difficulty is for instance to reconcile the successive claims of reintegration of phenomenal consciousness and first-person methods with the apparently incompatible idea of an anti-Cartesian drive, although Dennett provides one example of their compatibility.

Nevertheless, it offers the advantage of introducing sense and logic in an evolution rather disorienting at first sight, as well as of offering an interpretive grid to be used as a basis for investigating historical facts. In addition, it provides a way of understanding the present situation of cognitive science that suggests that the normative aspect of the problem of its foundations must henceforth be primarily reformulated in the following terms: How much can a theory of cognition be non-Cartesian, and how far beyond cognitive Cartesianism should cognitive science go? A proof that reconstructing the past of cognitive science is indispensable to the construction of its future.

References

Andler, D. (1992/2004) *Introduction aux sciences cognitives.* Paris: Gallimard.

Baars, B. J. (1986) *The Cognitive Revolution in Psychology.* New York: Guilford Press.

Bechtel, W., Abrahmsen, A., and Graham, G. (1998) The life of cognitive science. In: W. Bechtel, and G. Graham (eds.) *A Companion to Cognitive Science.* Oxford: Basil Blackwell.

Bennett, M. R., and Hacker, P. M. S. (2003) *Philosophical Foundations of Neuroscience.* Cambridge: Blackwell.

Bennett, M. R., and Hacker, P. M. S. (2008) *History of Cognitive Neuroscience.* Chichester, UK; Malden (MA): Wiley–Blackwell.

Boden, M. A. (2006) *Mind as Machine: A History of Cognitive Science.* Oxford: Oxford University Press.

Chomsky, N. (1957) *Syntactic Structures.* The Hague: Mouton.

Clarac, F., & Trenaux, J.-P. (2008). *Encyclopédie historique des neurosciences*. Bruxelles: de Boeck.

Clark, A. (1997) Being There: Putting Brain, Body and World Together. Cambridge (MA): MIT Press.

Damasio, A. R. (1994) Descartes' Error: Emotion, Reason, and the Human Brain. New York: G. P. Putnam.

Dennett, D. (1991) *Consciousness Explained.* Boston: Little Brown.
Dupuy, J. P. (2000/1994) *The Mechanization of the Mind: On the Origins of Cognitive Science.* Princeton (NJ): Princeton University Press.
Finger, S. (1994) Origins of Neuroscience: A History of Explorations into Brain Function. New York: Oxford University Press.
Fodor, J. (1987) Mental representation: An introduction. In: N. Rescher (ed.) *Scientific Inquiry in Philosophical Perspective.* New York: University Press of America.
Fodor, J. (2009) *LOT 2. The Language of Thought Revised.* Oxford: Oxford University Press.
Gardner, H. (1985) *The Cognitive Revolution.* Cambridge: Harvard University Press.
Lachman, R., Lachman, J., and Butterfield, E. (1979) *Cognitive Psychology and Information Processing.* Hillsdale (NJ): Lawrence Erlbaum.
McCorduck, P. (1979) *Machines Who Think.* San Francisco: Freeman & Company.
Miller, G. (1956) The magical number seven. *The Psychological Review* 63, 81–97.
Miller, G. (1960) Plans and the Structure of Behavior. New York: Holt.
Miller, G. (1979) A very personal history: MIT Center for Cognitive Science. *Occasional Papers* No. 1.
Newell, A., and Simon, H. A. (1972) *Human Problem Solving.* Englewood Cliffs (NJ): Prentice-Hall.
Newell, A. (1956) The logic theory machine. *IRE Transactions on Information Theory* 3, 61–79.
Newmeyer, F. J. (1986) *Linguistic Theory in America.* 2nd ed. Orlando: Academic Press.
Newmeyer, F. J. (1996) *Generative Linguistics: A Historical Perspective.* London & New York: Routledge.
Putnam, H. (1999) *The Threefold Cord: Mind, Body, and World.* New York: Columbia University Press.
Roy, J.-M. (1999) Cognitive turn and linguistic turn. *Proceedings of Twentieth World Congress of Philosophy.* http://www.bu.edu/wcp/MainCogn.htm. Boulder: Philosophy Documentation Center.
Roy, J.-M. (2010) The foundational crisis of cognitive science: Challenging the emergentist challenge. *Revista de Filosofia Aurora* 22 (30: Janeiro a Junho). Editora Champagnat. Curitiba, Brasil.
Searle, J. R. (1992) *The Rediscovery of the Mind.* Cambridge (MA): MIT Press.
Tête, A., and Pélissier, A. (1995) *Sciences Cognitives: Textes Fondateurs (1943–1950).* Paris: Presses Universitaires de France.
Triclot, M. (2008) Le moment cybernétique: la constitution de la notion d'information. Seyssel: Champs Vallons.
Wheeler, M. (2005) Reconstructing the Cognitive World: The Next Step. Cambridge (MA): MIT Press.

ISSUES AND ACTORS ON THE HISTORICAL SCENE

THIRTY YEARS OF COGNITIVE STUDIES OF CATEGORIZATION: WHAT IS BEHIND THE REPORTED PROGRESS?

Lilia Gurova

In the middle of the 1980s, the study of categorization was broadly viewed as one of the great success stories of the newly emerged field of cognitive science (see Gardner 1985). Two claims played an essential role in the standard account of this story:

1) The research on categorization led to the end of the domination of a more than two thousand-year-old view of concepts and categorization.
2) The disavowal of the old view of concepts and categorization became possible due to the contribution of three different disciplines – psychology, anthropology and philosophy –, a fact which has been seen since then as a clear demonstration of the advantages of the interdisciplinary approach to cognition.

H. Gardner wrote his book some thirty years after the official birth of cognitive science. The next thirty-year period is about to expire soon, and it is reasonable to ask what the continuation of the success story has been. Has the research on categorization made a significant progress since the time Gardner was so highly appreciated and what has made possible the alleged progress? The main aim of the present paper is to suggest answers to these questions. Before going to the main exposition, however, some terminological preliminaries have to be introduced.

Terminological preliminaries

By *category* by and large the cognitive scientists mean "a number of objects that are considered equivalent" (Rosch 1978: 191).[1] Categories are designated by *names* (e.g., *fish, table, stone),* or *combinations of names* (e.g., *pet fish, round table, philosophical stone).* Most cognitive scientists have assumed that *concepts* are the mental counterparts of categories (e.g., the concept of *dog* is the mental counterpart of the category *dog*). It is taken for granted as well that categories are organized in *category systems.* Some category systems called *taxonomies* play a special role in human cognition. The categories which form *a taxonomy "*are related to one another by means of class inclusion" (Rosch 1978: 191). For example, the categories

[1] The page numbers of all citations of (Rosch, 1978) refer to the reprint of Rosch's paper in Margolis and Laurence (1999).

bulldog, dog, mammal, and *animal* form a taxonomy insofar as all members of the category *bulldog* are *dogs,* all *dogs* are *mammals,* all *mammals* are *animals,* etc. Taxonomies have *levels* which are often called *levels of abstraction.* The more inclusive the given category is with respect to the other categories within the same taxonomy, the higher its level of abstraction is. By *categorization,* cognitive scientists mean the process of category formation in culture, as well as the identification of some entity X as a member of the category Y.

Categorization research today: Some common frustrations

Some important facts about the situation in the field today do not easily fit in a happy-end scenario. These facts even seem to cast a shadow on what has been earlier appreciated as a great achievement.

First, the old view of concepts, often labeled "the classical view" (see Smith and Medin 1981), and represented as the view stating that category membership is determined by a list of defining features, has not turned to be overthrown for ever. The view has revived under the name of rule-based categorization (see Erikson and Kruschke 1998), and decisive evidence has been obtained that most people do believe that the categories which they deal with possess defining features even if they are not able to list these features (Brooks 1978), and that under certain conditions people do rely on rules in categorization (Kloos and Sloutsky 2008).

Second, not one of the various versions of the suggested alternatives of the classical view (the prototype view, the exemplar view, the theory view, and the different hybrid versions of all these views)[2] has succeeded in getting support by the majority of cognitive scientists because none of these alternative views has been proven to fit well to all available experimental data.

Third, it is not a secret that none of the existing views on categorization really implies the so-called basic level effects which point to the existence of a privileged level within a taxonomic category system. Murphy (2002) may be the only one who has dared to mention this non-secret explicitly saying that in the best case, the existing theories and models of categorization are compatible with basic-levelness but do not predict it. At first glance, this fact does not seem to be a big trouble in itself. However, if we recall that the basic level effects have been broadly recognized as one of the most important experimental discoveries

[2] Those who are familiar with the main theories of categorization which have been discussed in the past thirty years may skip this footnote. Very briefly, the main statements of the suggested alternatives to the classical view of categorization are the following: According to *the prototype view*, also called *probabilistic view* (see Smith and Medin, 1981), categories are represented by (descriptions of) their best examples, and the category membership depends on the degree of similarity between the categorized item and the alleged best example (the category prototype). It is easy to see that the prototype view predicts fuzzy category boundaries insofar as in some border cases people may disagree whether a particular item is similar enough to the prototype. According to *the exemplar view* (Smith and Medin, 1981), categories do not have summary representations (neither in the form of a set of defining features, nor in the form of a prototype). They are represented instead by sets of exemplars. According to this view, categorization consists in checking whether the categorized item is similar enough to any of the exemplars which represent the given category. *The theory view* of categorization (Murphy and Medin, 1985) states that category membership is determined by a theory. The most important implication of this view is that categorization is not necessarily similarity-based. This means that very dissimilar objects may belong to the same category if there is a theory which tells us in what respect these objects should be considered the same kind of objects.

of the early research on categorization, and that an essential part of the criticism against the old classical view of categorization has been built on the fact that the classical view does not predict typicality effects (Smith, and Medin 1981), it becomes evident that the alternatives of the classical view which cannot predict basic-levelness are not in a much better position.

At the same time, in one of the last extensive reviews of the research on concepts and categorization, Murphy (2002) stated clearly that despite the problems which the current theories of categorization have, there has been real progress in the field: since 1980s "the field has greatly expanded", he says, "covering many topics that did not really exist then" and "much more has been learned" about the topics that were in the focus in those early years of categorization research (Murphy 2002: 8). One can reasonably doubt, however, whether the expansion of a research field should be taken as a clear sign of progress. Simply adding new topics to the research agenda does not mean that the initial problems have received a satisfactory solution. On the contrary, very often such problem shifts take place when the problems which a research program has initially stated as central turn out to be not directly tractable (Lakatos 1970).

The uneasy facts listed above about the present results of the research on categorization seem to suggest that no significant advance in our theoretical understanding of human categorization has been realized since the 1980s, and that the only progress that has been made in the field consists in a number of empirical findings and a few interesting, but only locally valid generalizations drawn on these findings.[3] This paper will present a different perspective on the history of categorization research. This perspective reveals that the early research on categorization gave rise to important theoretical inventions, which continue to inspire and to give a meaning to many interesting empirical results. In order to recognize this theoretical advance, however, one should go beyond the standard account of the early success story, which is silent about the mentioned theoretical insights.

In what follows, the standard account of the recent history of categorization research will be presented first and the main flaws of this standard account will be outlined.[4] After that, the line of research will be described for which the claim of this paper is that it has led to a genuine progress in understanding the phenomena of categorization.

The standard account of the history of categorization research and its discontents

According to the standard account, see, for example, Smith and Medin (1981), since the time of Aristotle, a particular view of concepts has dominated the Western mind. This is the view that concepts refer to groups of entities which possess the same defining features. According to this view, for example, we identify a given object as a *dog* because that object possesses all the features that define the category of dogs. This defining features view, for which Smith

[3] Murphy (2002), for example, has suggested a similar account of the progress made in the categorization research area.

[4] I have already criticized the standard account of the history of categorization research in Gurova (2003) from a different perspective. One of the central claims of my 2003 paper was that the standard account is a source of unsound criticism of what has been called "the classical view" of concepts and categorization. Different aspects of the standard account are at stake in the present paper.

and Medin launched the name "classical", has always had its critics, the standard story holds, but it was seriously challenged only when some important experimental results were obtained which turned out to be not implied and, therefore, not explainable by the classical view. In the 1970s, it was reliably established, in the first place, that people tend to treat some category members as more typical than others, which is difficult to explain on the assumption of the classical view that all category members possess the same defining features. As a result, some alternatives to the classical view emerged, but each of them only partially succeeded to cope with the wealth of available empirical data. This fact has been most often explained by the assumption that probably different types of category representations interact and produce what we recognize as "categorization" (Kruschke 2003). This explanation does not exclude the possibility for a more general or encompassing theory of categorization, but given the context of the classical view and its alternatives, nobody seems to be able to guess what this encompassing theory has to look like.

What does this standard representation of the recent history of categorization research miss?

First, the standard account outlined above emphasizes the importance of only one part of the empirical findings obtained in the 1970s by E. Rosch and her collaborators. These are the typicality effects mentioned above. Rosch herself, however, has stressed much more the importance of the so-called basic level effects or the fact that people have a preferred level of categorization within a given conceptual taxonomy (Rosch et al. 1976; Rosch 1978). A manifestation of the existence of a basic level of categorization is, for example, the experimentally confirmed fact that given the taxonomy *bulldog, dog, mammal, animal,* most subjects categorize an object as a *dog* rather than as an *animal* or *bulldog,* although they know that dogs are animals, and even if they know that this particular dog is a bulldog. However, in the literature reviews written in the perspective of the standard account (see Smith and Medin 1981), the basic level effects have been only mentioned in passing. This neglect is understandable. The role of basic level effects seems insignificant for the heroic story about the demise of the classical view of categorization and the rise of its "more natural" (Gardner 1985) alternatives because these effects do not constitute evidence against the classical view. But as Murphy (2002) stressed, basic level effects have been a genuine discovery, and any story which fails to reveal the significance of such a genuine discovery should be treated with reservation.

Second, in her writings summarizing the chief achievements of her and her collaborators' research, Rosch (1978) herself does not point to the disavowal of the classical view as to the most glorious result of her studies. The central claims she stressed in her argument supporting the claim that human categorization is the result of two main principles of categorization were the principle of cognitive economy, and the principle of perceived world structure. These two principles, according to Rosch, underlie any category system (Rosch 1978), and imply both typicality effects and basic level effects. Why, however, has Rosch's own theoretical account of typicality and basic-levelness been ignored and superseded by the obviously dead-end standard account? The following historical reconstruction will provide an answer to this question.

The success story of categorization research revisited

As it was mentioned above, G. Murphy (2002) claimed that basic level effects were a genuine discovery, made by Rosch and her collaborators. This is, however, not entirely true. It was the psychologist Roger Brown (1958), who first noticed that people prefer to use a particular level of categorization in speech: i.e., when they see a dog, they prefer to call it *dog* instead of *bulldog* or *animal.* In 1972, B. Berlin and his collaborators added to this observation a new one: that in the case of folkbiological categorization, the same taxonomic level has been preferred by representatives of different cultures. They speculated that this is the level which corresponds to natural groups of organisms, and that these natural groups can be easily identified as such because their members possess some common and correlated features. According to Berlin and his fellow-ethnobiologists, the natural groups can be found at the level of biological genus.

Rosch, however, was not satisfied with the ethnobiological explanation of the privileged level of categorization. Her previous involvement in cross-cultural studies of the categorization of colors convinced her that even in cases where there are no natural groupings (like the color spectrum, which is physically continuous), the subjects belonging to different cultures tend to form the same color categories. From cases like this she concluded that there must be some psychological (in addition to the physical) determination of the process of categorization. Thus, she arrived at the following two principles of categorization (Rosch 1978: 190):

1) The principle of cognitive economy: "The task of category systems is to provide maximum information with the least cognitive effort".
2) The principle of perceived world structure: "The perceived world comes as structured information rather than as arbitrary or unpredictable attributes".

According to Rosch, these two principles of categorization have implications for both the vertical and the horizontal dimension of category systems:

> The implication of the two principles of categorization for the vertical dimension is that not all possible levels of categorization are equally good or useful; rather, the most basic level of categorization will be the most inclusive (abstract) level at which the categories can mirror the structure of attributes perceived in the world. The implication of the principles of categorization for the horizontal dimension is that to increase the distinctiveness and flexibility of categories, categories tend to become defined in terms of prototypes or prototypical instances that contain the attributes most representative of items inside and least representative of items outside the category. (Rosch 1978: 191)

Thus, what Rosch saw as her greatest theoretical achievement was the discovery of the principles that explain the two most important experimental findings in the categorization research: basic level effects and typicality effects which have been regarded so far as separate phenomena. Rosch suggested four operational definitions of the supposed basic level of categorization in taxonomic systems. According to these four definitions, the basic level in a given taxonomic category system is the most inclusive level at which

(D1) the members of the category possess a significant number of common attributes,

(D2) the subjects use the same motor programs to deal with the category members,
(D3) the members of the category have a similar shape,
(D4) the subjects can form an average image to represent the category, and they can use this image in categorization.

The results of experiments done by Rosch and her collaborators and presented in Rosch et al. (1976), and Rosch (1978) revealed that all four definitions point to the same basic level in different taxonomic systems. This result was broadly accepted as an evidence for the reality of the basic level of categorization. There was a problem, however. Namely, in the case of folkbiological categorization, the experimental results obtained by Rosch and her fellows revealed as basic the level of "life-form" (e.g., *tree, fish* etc.) and not the level of genus (e.g., *maple, salmon* etc.), which was pointed as the privileged level of categorization by ethnobiologists.[5] Rosch herself tried to explain this discrepancy of the results by referring to the limitations of ethnobiological methodology, which, according to her, asserted the existence of "natural groupings" at the level of genus relying exclusively on the existence of few correlated attributes shared by the members of these groupings (Rosch 1976: 386).

Most authors after Rosch, however, regarded the discrepancy as a demonstration of a context effect and the main context factor, which they stressed was the level of expertise: "urban dwellers treat the life form – in this case, *tree* – as basic, rather than the genus, such as *maple* or *elm* (Dougherty 1978; Rosch et al. 1976), presumably because of lesser amounts of interaction with the natural environment" (Murphy 2002: 211).[6]

However, the context effect explanation of the instability of results about the basic level confronts us with a serious problem. If the basic-levelness depends on human experience, then in what sense (and to what extent) may one assert the existence of psychological (rather than cultural, or social) determination of human categorization? Rosch herself was aware that context sensitivity is a "problematic issue" for her theory of categorization although she tried to depreciate the possible threats of the context effects by saying that "both basic levels and prototypes are, in a sense, theories about context itself" (Rosch 1978: 202). She, however, did not say much about in what sense basic levels and prototypes are theories about context.

Rosch's principles of categorization were broadly ignored in the following years mainly because most researchers in the field became occupied with the question of how people represent categories in their minds, and how they use these category representations in different categorization tasks. As Rosch herself stressed in her 1978 paper, her theory was not about that, it was about the formation of categories in culture. The standard account of the categorization research has been based mostly on this post-Roschean research inspired by the question of what kind of category representations can cope with the empirical facts about human categorization. Thus, it focused on the rival views of category representations, and the extent to which the different views have been confirmed by the evidence available. That explains

[5] It should be stressed here that the main folkbiological taxonomical ranks – folk kingdom, life form, generic-species (or folk generic) – only partially coincide with scientific biological ranks (kingdom, division, class, order, family, genus, species). In a particular taxonomy system, the folk generic, for example, corresponds either to the biological species, or to the biological genus. The folkbiological life-form is even more loosely connected to the scientific biological ranks, which are higher than the biological genus but lower than the biological kingdom (see Medin and Atran 2004).

[6] See also (Tanaka and Taylor 1991).

why the standard account has omitted Rosch's early attempt to reveal the principles underlying all category systems and to provide a common explanation for both typicality and basic-levelness: These principles turned to be useless for the research which put in the focus the questions of how people represent categories in their minds and how they use these alleged mental representations in everyday categorization and reasoning.

A minority of researchers, however, did not forget about the notice that the discovery of the basic level effects is indeed one of the most important empirical findings (if not the most important one) in the recent history of categorization research. Thus, they continued to look for an explanation of how is it possible for the basic level to be so stable (universal across cultures) and sensitive to context effects at the same time.

The most significant advance in the search for the solution of this problem was made possible again thanks to the cooperation of anthropologists and psychologists. The leading figures in the new groundbreaking research were the anthropologist Scott Atran and the psychologists Douglas Medin (well-known as well for his contribution to the establishment of the post-Roschean agenda of the categorization research, and the standard account of the early success story of this research).

Medin and Atran began the work on their common project in the early 1990s. What brought them together was the idea which both of them believed in that cross-cultural studies are the only way which might lead to the discovery of what is truly universal in human cognition and what is susceptible to cultural variations. At that time, Atran had already published his book *Cognitive Foundations of Natural History* in which he launched the idea that

> children, tribal people, modern layfolk – even scientists in their nonworking hours – readily partition the ordinary range of human experience accordingly to cognitive domains that are pretty much the same across cultures" and which are the product of "millions of years of biological and cognitive evolution. (Atran 1990: x)

Biological cognition is one of these domains which turn to be especially good for cross-cultural studies because the biological world provides a natural metrics for comparing the systems of biological knowledge developed in different cultures. Both Atran and Medin agreed with what ethnobiologists had already discovered, namely that "striking cross-cultural similarities suggest a small number of organizing principles that universally define systems of folkbiological classifications" (Atran and Medin 2008: 33). Thus, they decided to undertake a large scale project, convinced that cross-cultural studies, combining the ecological validity of anthropological field research with the precision of psychological experimentation and informed by the hypothesis that biological cognition constitutes a "distinct module of mind that is associated with universal patterns of categorization and reasoning" will provide "a new perspective on a range of fundamental issues in cognition" including "relative contributions of universal versus culturally specific processes to people's conceptions of biological kinds" (Atran and Medin 2008: 15).

Previous research has revealed two candidates for universal organizing principles: universal structure of folkbiological classification, and the universal notion of biological essence for which it has been discovered that in all cultures biological essence is associated with the biological kinds at the generic-species level (also called folk generic), which corresponds to the joint ranks of species and genus in the scientific biological classification. It is the associa-

tion of biological essence with a particular level of folkbiological classification that makes this level privileged or "basic" if we decide to follow the terminology introduced by Rosch. And here Atran and Medin face the paradox which has already been mentioned in this paper: if the basic level is psychologically determined by universal cognitive capacities which have been favored by evolution for millions of years, this basic level should be the same across cultures and unmovable. Rosch's experiments, which have been replicated many times in the following years, reveal that in populations which have grown up under conditions of limited access to the living world the basic level shifts to the upper "life-form" folkbiological rank. Atran and Medin also made the striking discovery that only the recognition strategy moves to the upper level (i.e., subjects with normal exposure to the biological world will recognize the given object as a *maple,* while subjects with limited biological experience will recognize it as a *tree).* The reasoning strategy remains intact. For example, when the subjects are given the following two premises:

P1: All Xs are Y.
P2: All Xs have the property P.

And after that they are asked the question:

Q: How many of Y have P, too? (With possible answers: all, few, none)

Both biologically experienced people (farmers, botany experts, living in a close contact with nature Indian tribes etc.), and biologically impoverished inhabitants of big cities agree that most Y have P too, if Y is at the generic-species level, and again both groups agree that rather few Y may have P when Y is a biological group of a rank higher than the generic species.

Why does experience not matter in such inferential tasks? Medin and Atran's answer is that it is because of the presumption of essence associated with the generic species level (1999, 2004). Inductive inferences are directly based on this presumption, which, according to their hypothesis about folkbiological system is in-built. The recognition strategy is more movable because it depends crucially on the available knowledge about the perceived world.

Thus again, theoretical speculations were proved successful in resolving a more than 30-year-old puzzle. And this was not an ad-hoc explanation but rather one derived from principles which have been justified independently by converging evidence obtained from different sources.[7]

Some final remarks

The cognitive studies of categorization do have their success story. The cross-disciplinary cooperation did contribute to this success story at least twice. First time, it happened in the 1960s and the 1970s when anthropologists noticed that a privileged basic level exists in human taxonomical category systems. After that, their observation made possible the

[7] See Medin and Atran (2004), and Atran (2001) for discussion on the evidence for conceptual modularity.

psychological discovery of the basic level effects, which have been proven to penetrate almost all cognitive processes from simple categorization to higher level inductive reasoning. The latest demonstration of the advantages of the cross-disciplinary cooperation in categorization research are the results of the cross-cultural project of Medin and Atran, which was launched in the early 1990s, and a summary of their results was published in 2008. One of the greatest achievements of this joint project is the solution of the paradox of basic-levelness: how is the basic level universal across cultures and sensitive to context effects at the same time? The answer to this question which Medin and Atran suggested is the following: evolutionary determined and universal across culture basic level in human folkbiological taxonomic systems does exist. Its manifestations are context sensitive in tasks like visual recognition for which previous experience/knowledge matters. However, the basic level effects appearance is quite uniform and stable in reasoning tasks for which previous experience and knowledge do not seem important.

The cross-disciplinary cooperation, however, is only one of the factors contributing to the success story. The other factor which has been less recognized so far is the role of the theoretical insights based on evolutionary considerations. These theoretical insights include Rosch's principles of categorization which have allowed for a common explanation of the seemingly diverse typicality effects and basic level effects. No less important is the theoretical insight underlying the speculations of Medin and Atran about the existence of evolutionary determined conceptual modules (and of a folkbiological module in particular), which suggests a quite successful explanation of the earlier controversial results about basic-levelness. It is important to stress the role of evolutionary considerations in the success story of categorization research because evolutionary considerations have not often been favored by cognitive scientists. On the contrary, eminent cognitivists (to mention only Chomsky, Thagard, Fodor) are famous for their criticism of different evolutionary projects concerning cognition. The perspective presented here on the history of categorization research reveals that Dobzhansky's famous slogan that "nothing in biology makes sense except in the light of evolution" (Dobzhansky 1973) is to be taken seriously in respect to cognition, too. At least because to this moment there are no better candidates for a common theoretical framework which can serve as a unifying explanatory background for such a great variety of effects concerning categorization.

References

Atran, S. (1990) *Cognitive Foundations of Natural History.* Cambridge: Cambridge University Press.

Atran, S., and Medin, D. (2008) *The Native Mind and the Cultural Construction of Nature.* Cambridge (MA): MIT Press.

Berlin, B. (1972) Speculations on the growth of ethnobotanical nomenclature. *Language in Society* 1, 51–86.

Brooks, L. R. (1978) Nonanalytic concept formation and memory for instances. In: E. Rosch, and B. B. Lloyd (eds.) *Cognition and Categorization.* Hillsdale (NJ): Lawrence Erlbaum.

Brown, R. (1958) How shall a thing be called? *Psychological Review* 65, 14–21.

Dobzhansky, T. (1973) Nothing in biology makes sense except in the light of evolution. *American Biology Teacher* 35, 125–129.

Dougherty, J. W. D. (1978) Salience and relativity in classification. *American Ethnologist* 5, 66–80.

Gardner, H. (1985) The Mind's New Science: A History of the Cognitive Revolution. New York: Basic Books.

Gurova, L. (2003) Philosophy of science meets cognitive science: The categorization debate. *Boston Studies in the Philosophy of Science.* Vol. 236. Dordrecht: Kluwer, 141–162.

Erikson, M. A., and Kruchke, J. K. (1998) Rules and exemplars in category learning. *Journal of Experimental Psychology: General* 127, 107–140.

Kloos, H., and Sloutsky, V. (2008) What's behind different kinds of kinds: Effects of statistical density on learning and representation of categories. *Journal of Experimental Psychology: General* 137, 52–72.

Kruschke, J. (2003) Concept learning and categorization models. In: L. Nadel (ed.) (2003) *Encyclopedia of Cognitive Science.* London: Macmillan.

Lakatos, I. (1970) Falsification and the methodology of scientific research programmes. In: I. Lakatos, and A. E. Musgrave (eds.) *Criticism and the Growth of Knowledge.* Cambridge: Cambridge University Press.

Margolis, E., and Laurence, S. (eds.) (1999) *Concepts. Core Readings.* Cambridge (MA): MIT Press.

Medin, D., and Atran, S. (1999) *Folkbiology.* Cambridge (MA): MIT Press.

Medin, D., and Atran, S. (2004) The native mind: Biological categorization and reasoning in development and across cultures. *Psychological Review* 111, 960–983.

Murphy, G. (2002) *The Big Book of Concepts.* Cambridge (MA): MIT Press.

Murphy, D., and Medin, D. (1985) The role of theories in conceptual coherence. *Psychological Review* 92, 289–316.

Rosch, E., Mervis, K., Gray, W., Johnson, D., and Boyes-Braem, P. (1976) Basic objects in natural categories. *Cognitive Psychology* 8, 382–439.

Rosch, E. (1978) Principles of categorization. In: E. Rosch, and B. B. Lloyd (eds.) *Cognition and Categorization.* Hillsdale (NJ): Lawrence Erlbaum. (Reprinted in: E. Margolis, and S. Laurence (eds.) (1999) *Concepts. Core Readings.* Cambridge (MA): MIT Press.)

Smith, E., and Medin, D. (1981) *Categories and Concepts.* Cambridge (MA): Harvard University Press.

Tanaka, J. W., and Taylor, M. F. (1991) Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology* 15, 121–149.

ANIMAL MEMORY AND THE ORIGINS OF MIND: THE CONCEPTION OF LAJOS KARDOS, A HUNGARIAN COMPARATIVE PSYCHOLOGIST[1]

Csaba Pléh

The main goal of this paper is to present the relevance of Lajos Kardos's work to the history of the cognitive movement at large. Kardos was a Hungarian comparative psychologist in the mid-20th century. Kardos will be presented as an author whose work span from the Gestalt perceptual tradition to modern cognitive psychology. Kardos represented three general ideas that have interesting parallels in contemporary cognitive theory. The first is the need for a mathematical treatment of contextual effects in constancy phenomena as related to object perception. The second topic is his theory of animal memory proposing that mammalian memory achievement depends on image-based processes, not unlike the cognitive maps of Tolman (1948). As a third achievement, Kardos also proposed an even more general theory about the genesis of representative processes, where representation is treated as a key to mental life, and it is reduced to the mapping of corollary informations about the environment. In a peculiar way, representation is an exapted function in the sense of Gould and Vrba (1982), though Kardos himself of course never used this term that later had evolutionary considerations.

Lajos (Ludwig) Kardos (1899–1985)

A life full of challenges

Lajos (Ludwig) Kardos (1899–1985) was both the mentor and the savior of Hungarian experimental psychology in the 1950s and 1960s acting as a chair of psychology at the Budapest Eötvös University between 1947 and 1972, at a time when psychology was less

[1] Earlier versions of this paper were kindly read and commented on by two leading vision researchers familiar with the perceptual work of Kardos, Allan Gilchrist of Rutgers University and Deján Todorovic of the University of Belgrade helped my presentation with their comments. I wish to thank their patience and help. In preparing the final version of this chapter tha author was enjoying a scholarship at the Collegium de Lyon, ENS Lyon and a support within the framework of the TÁMOP-4.2.2.C-11/1/KONV-2012-0008 (Social Renewal Operative Program) project titled *The application of ICT in learning and knowledge acquisition: Research and Training Program Development in Human Performance Technology.* Said project was implemented by the support of the European Union and the co-financing of the European Social Fund.

Table 1: Kardos as the chief mentor of a generation of psychologists (data after Bodor, Pléh, and Lányi 1998; Pléh, Bodor, and Lányi 1998)

Psychologist–scholar	References
Kardos, Lajos (Ludwig)	86
Mérei, Ferenc	45
Harkai Schiller, Pál	33
Szondi, Lipót	29
Ferenczi, Sándor	22
Hermann, Imre	18
Várkonyi, Hildebrand	18
Radnai, Béla	17
Benedek, István	13
Gímesné Hajdú, Lilly	12
Lénárd, Ferenc	11

than welcome as a discipline (Pléh 2008). Table 1 shows his importance in autobiographic references. The table lists references to Hungarian psychologists in autobiographies of psychologists written on their own in the 1960s and 1970s, and during the late 1990s. As can be seen from the numerous references, Kardos was the chief mentor and reference figure of the age; most of the others mentioned are psychoanalysts, and Ferenc Mérei is an early social psychologist.

Kardos achieved this status after an exodus full of typical features of the life of East European intellectuals, becoming established as an academic leader in Hungary only in his late 40s.

He was born in Rákospalota, then a suburb of Budapest (nowadays a part of the big city), on 14th December, 1899, in a 'small bourgeois' Jewish family, and died in London, on 12th July, 1985, while visiting his emigrant daughter there. Kardos became part of the Jewish exodus in the 1920s due to *numerus clausus,* and he started and finished his university studies in Vienna. *Numerus clausus* was the first practically anti-Semitic law in Hungary that tried to limit the proportion of Jewish students to 6%, which was the proportion of Jews in the general population of Hungary at that time, while the actual rate of Jewish students was 25–40% in the 1910s in different faculties (Kovács 1994). This law was responsible for many would-be Hungarian Jewish intellectuals to make them study abroad. Kardos studied both medicine and mathematics at the University of Vienna, obtaining his medical degree in 1925. But the real turning event of his life was that in the 1920s he became a student of Karl Bühler (Kardos 1984a; Pléh 1985; Murányi 1985). After defending his thesis, he published it in Nazi Germany (Kardos 1934), with some benevolent lying from Bühler as Kardos mentioned in an interview (Pléh 1985). As Dejan Todorovic (2010) pointed out to me, "[t]he book was dedicated to Karl Bühler, and in the preface Kardos extends thanks to various people including Bühler, Spearman, Woodworth, Brunswik, Heider, MacLeod, and especially Koffka, for 'long and

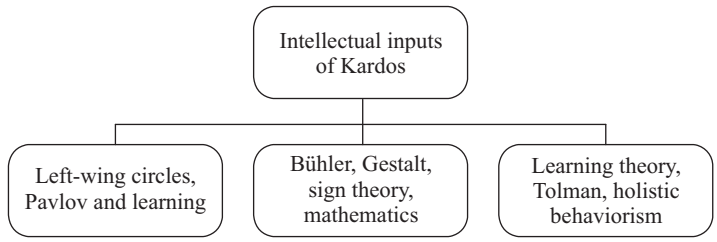


Figure 1. Young Kardos's intellectual inputs

deep discussions'. This shows that he was in communication with leading researchers in the field at that time."

Kardos left for the US in the 1930s as a Rockefeller Scholar and taught at some American colleges, among them at Wells College, near Ithaca (N.Y.), a private woman's college at the time.

In a rather irrational way, but not unlike the other Hungarian mentor of his generation, Mérei, who returned home from France, Kardos returned to Europe and Hungary in the late 1930s, and became part of the circle around Lipót Szondi, an influential biologically oriented depth psychologist. Kardos was trying to become a practical clinician.

Surviving the war, from 1947 he became the founding chair of the Department of Psychology (later Department of General Psychology) at Loránd Eötvös University, Budapest. This was the second time the Department was established there. In 1918–1919, the department already existed under the leadership of Géza Révész (Pléh 2008).

After its reconstituted rights, Kardos was nominated and elected to be the first new president of the Hungarian (Scientific) Psychological Society in 1960.

Among the honors, Kardos received a doctor honoris causa title from the University of Padua, Italy, where he performed many of his experimental works on animal learning and memory, and a membership in the Hungarian Academy of Sciences shortly before his death in 1985.

As a researcher, Kardos started to work under the influence of three innovative movements in the 1930s, as summarized in Figure 1.

Already as a high school student in Budapest, Kardos regularly went to public and semi-private meetings of the most educated left-wing liberal intellectual circles, where his openness towards social ideas, and modern science developed. In Vienna, through Karl Bühler (1913, 1927, 1934), Kardos became involved in three approaches to psychology. The idea of Gestalt organization, the framework of treating the mind in sign theoretical control terms, and relating this to animal behavior all originated in Bühler's work. However, Kardos's separate outstanding contribution was the idea to deal with perception in mathematical terms.

The few years he spent in the US exposed him to learning theory, where Tolman's (1932) holistic approach was more appealing to the Gestalt-trained Kardos than the molecular attitude of behaviorist scholars like Hull (1934).

Young Kardos's perceptual research

Starting as a student of Karl Bühler, Kardos worked on constancy phenomena (Brunswik and Kardos 1929; Kardos 1930), and he became well-known through his monograph on the role of

shadows and lightness constancy in object perception (Kardos 1934). He was among the first ones among perceptual psychologists to combine the attitudes of careful experimentation with courageous mathematical modeling, basically claiming that constancy can be rendered with a mathematical model comparing the light input from a surface with that of the neighborhood (Kardos 1934, 1935).

According to Kardos, color and lightness constancy phenomena are a key to object perception. Constancy itself can be rendered with a mathematical model comparing the light input from a surface with the average light coming from the neighborhood. As Alan Gilchrist (2010) pointed out to me in personal correspondence, "[t]he idea that lightness depends on a comparison of target and surrounding luminance was, I think, widely accepted among at least Gelb, Koffka, and others. …his idea of "neighborhood" was much more concrete. It was not a matter of distance from a target surface, but rather a frame of reference. He used the terms relevant and foreign "field" as in the field of illumination. Furthermore, it was not the kind of vague idea others had. He defined how a field is segregated within a complex image by two factors: penumbra, and depth boundaries (corners and occlusion boundaries)" (Gilchrist 2010).

Kardos's treatment of constancies is a rather striking combination of phenomenological analysis, careful experimentation about contextual effects, and an innovative application of higher mathematics. In his phenomenological analysis, there is careful consideration of notions like object, field, sign, and the like. Phenomenology for Kardos is by far not a license for loose talk. Rather, it is a combination of conceptual analysis and the presentation of primary experiences.

> In the natural, lay attitude directed towards „object properties" vision provides a phenomenal field in which there is no real articulation between shadows and parts without a shadow similar to a figure–ground organization. (Kardos 1934: 23)

In the 1984 Hungarian translation of his 1934 monograph, he felt pity for his missing of cybernetic notions that would turn the phenomenological language into a more mathematically neutral idiom:

> How much easier would have been my task (in 1934) had I available the conceptual apparatus of present day information theory and cybernetics! [..] How easier would it have been to state that our color experiences are informations about some optical aspects of objects, and to state that stimuli work as information channels are characterized by noise. (Kardos 1984b: 13)

Going along with this line of reasoning, Kardos (1964) was among the first to emphasize within the Hungarian context – where cybernetics was a dubious bourgeois science in the earlier Stalinist times – that cybernetic notions such as regulation, control, feedback, and the like are key to understand what he referred to as 'correction systems' in perception.

Kardos, along with the Viennese ideas of representative design (Brunswik 1934, 1956) used contextual experimentation with elaborate situations, which represent something of the world. This is a characteristic feature of his later comparative studies as well. In his famous disk and shadow experiments, he was looking for changes of the experience of the disk as a

function of seeing the outline of the shadows as Figure 2 shows: *grey disk* vs. *white disk in a shadow.*

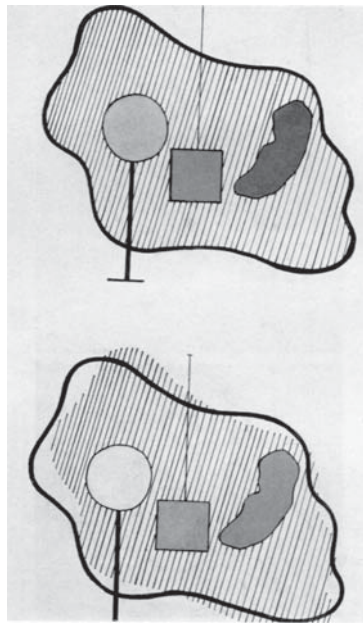


Figure 2. The changing shadow situation. Subjects have to report phenomenal changes in the experience of the circle disk. The lower part shows when by changing the shadow outlines they are made to realize that the circle is in fact in a shadow (Kardos 1934)

Kardos in his mathematical model was treating constancies as an example of co-determination. As Gilchrist (2004) characterizes him, "[h]is most important theoretical contribution was the concept of co-determination. Kardos argued that the lightness of an object is never computed exclusively relative to its own frame of reference, but rather shows an influence from foreign, or adjacent fields of illumination as well."

In Kardos's mathematical model, this was shown as a set of differential and integral equations relating luminance from the object and the surrounding field.

$$\frac{\frac{a}{L'}\int\int_{\xi}\frac{\Phi(X_0X)dxdy}{\sqrt{(x-\xi)^2+(x-\xi)^2}}}{1}$$

Kardos is still present even in contemporary psychology with his perceptual research both due to his outstanding experimental skills (these were already identified in his own time, see Crannell 1948), and due to his mathematical modeling of contextual effects.

Kardos proposed that the lightness of a target is co-determined, partly by its relevant field of illumination and partly in relation to what he called the foreign field of illumination. The Katz light/shadow arrangement can be used to illustrate the concepts of Kardos. For

> the target in bright illumination, the lighted region is the relevant field and the shadowed region is the foreign field. These roles are reversed for the target in shadow. It is an empirical fact that, except in special cases, the lightness value of a given target does lie somewhere between its lightness value when computed relative to the highest luminance in the relevant field and its lightness value when computed relative to the highest luminance in the foreign field. (Gilchrist and Annan Jr., 2002, see also Gilchrist 2006)

The importance of his combination of contextual experimentation with mathematical models of context effects was also emphasized by his Hungarian followers (Tánczos 1977).

Incidentally, beside his theoretical interest, Kardos's perceptual experimental work still shows up in motivating actual perceptual research today, dealing with illusion phenomena (Logvinenko, Petrini, and Maloney 2008), and with the role of reflectances in shape perception (Fleming, Torralba, and Adelson 2004).

In the 1960s, while Kardos moved to animal psychology in his experimental research, he did not entirely loose his theoretical interest towards perception. He made attempts to interpret the famous Innsbruck adaptation studies of perception performed by Ivo Kohler (1951, 1962, 1963). For Kardos, the essential aspect of these adaptation studies was again the 'regression towards the object': "[t]he final stage, the mental event covaries merely with a single earlier event, with the object" (Kardos 1965b: 15). The quote in fact comes from a talk held at the University of Innsbruck in 1963. Kardos was probably invited there by his junior colleague Ivo Kohler (1915–1985). Another treatment of the same experiments (Kardos 1966) is based on a talk at a conference of the *Kongress der Deutschen Gesellschaft für Psychologie*. Thus, the ideas of Kardos about how to combine a cybernetic way of talking with the traditional phenomenological attitude were in principle known to the German experimental psychology community in the 1960s.

In these talks, Kardos also prefigured his later attitude towards mental and neural organization:

> The neural event is a necessary and sufficient precondition for experience not due to its inherent characteristics but due to its role fulfilled in the integral psychophysical series of events. (Kardos 1966: 29)

According to Kardos, there is no analogy between phylogenesis and adaptation in these adaptation phenomena:

> The natural disturbances of the integral psychophysical series of events such as fluctuations of illumination, and retinal shiftings due to body movements are present at the birth of vision, and flow into its formation during hundreds of thousands of years; during phylogenesis there are no transformations in order to control for similar fluctuations. (Kardos 1966: 31)

From conditioning to animal memory

During the 1950s, psychology had a difficult time in Hungary. Psychology training was practically non-existent, and all psychology was classified under education in the Communist science management. Kardos was almost the single survivor of experimental psychology in

Hungary, with a handful of followers and associates like Ilona Barkóczi, Magda Marton, and Zsolt Tánczos. There were intensive Marxist campaigns against psychology, which made psychology discredited both as a profession and as a science (Pléh 2008). As Kardos himself recalled in an interview (Pléh 1985), he tried to convince one of his youth left-wing mentors from the Galilei Circle, Béla Fogarasi, then an influential leader of Communist party line philosophy at the university that some sort of natural science-minded psychology might still be possible. They agreed that learning was a crucial social topic where psychology can have a, say, helping education. That was the time of the intensive Pavlovization of psychology in the Soviet Union. Kardos started on a dangerous excursion: under the impact of his American travels and the new ideological needs of socialist Hungary, he tried to combine Pavlovian ideas on conditioning with learning theory in the American sense (Kardos 1960). This effort half a century later is a historical testimony about the difficulty of combining the open-ended experimental tradition with the cloze-minded ideological interpretation where the idolized Pavlov would have a prefabricated answer to all empirical and theoretical issues.

It is more important from our perspective, however, that Kardos initiated a long series of experimental studies on animal learning and memory in rodents from the 1950s on. On the theoretical level, he started from an analysis of the relationships between the "animal way of life" and mental organization. In this regard, he is a Gestaltist who was sensitized in the circle of Bühler (1934) to the ideas of early ethology emphasizing species-specific behavior and the different *Umwelts* of animals. Kardos interprets Köhler (1921) and pays tribute to the Marxist ideas regarding the relevance of labor in hominid evolution (Engels 1876). For Kardos, the essential difference in the way of life between mammals and apes is the *opposition between locomotion and manipulation.* In his own words, *„[t]he essential aspect of hominid evolution might be that the locomotory learning ability is replaced by manipulatory activity"* (Kardos 1959: 54; see also Kardos 1965a).

This topic remained crucial for Kardos for the rest of his life. Several years later, he performed some experimental work showing that manipulation-based learning in humans was easier than locomotion-based learning (Kardos, Barkóczi, and Kónya 1971).

Kardos's actual animal learning experiments were run through 30 years. They were sometimes done in pathetic circumstances (for a detailed description, see Barkóczi 1998). The studies started in Kardos's flat, which was in a city apartment building, then were performed in two buildings of Eötvös University, on the top floor in one, and in the basement in another building, and also at the University of Padua, in the Otology Clinic. The latter is related to the fact that some of the studies involved the surgery of the vestibular apparatus. Kardos was also invited to Padova by his fellow Gestaltist, Fabio Metelli (1907–1987), his regular visitor in Budapest.

The studies were initially done in *maze learning.* Kardos used ingenious techniques to show the peculiarities of animal memory tied to locomotory activity. Basically, Kardos showed that animal memory is place-tied, rats being unable to learn different targets being on the same place if the place was reached by different routes. Rodents have an image-like quasi perceptual memory that stores things together with their localizations.

His starting point was the idea that behavioral equivalence is crucial to learning. That is an idea again that goes back to the concept of behavioral equivalence claimed by his teacher Bühler (1927) in the framework of early continental ethology. The underlying sign-based equivalences in animal learning for Kardos are the following:

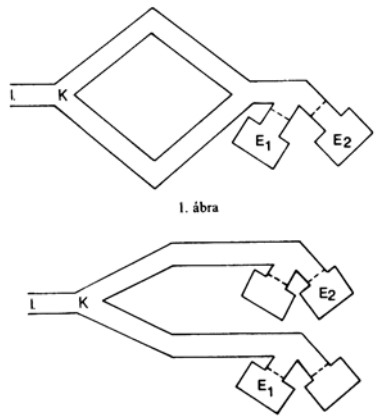


Figure 3. Two types of mazes where the animal has to choose on the basis of the pathway leading to the choice point (upper half), or when the animal stands at two different points (lower part) (Kardos and Barkóczi 1953)

The first studies along this line were his experiments on "aequiterminal routes"[2] (Kardos and Barkóczi 1953). Rats had to learn two slightly different types of mazes. The maze in the upper half of Figure 3 has two diverging routes leading to the same place. Which one is the positive goal box depends on the route taken by the animal. However, the animal stands on the same choice point when choosing according to the route passed. In the maze shown in the lower half, on the other hand, the two diverging routes do not meet again; the animal is at two different choice points after running along the upper and lower pathways.

The upper maze version is impossible to learn for rats. Thus, rats are not able to learn the distinction that if you came from left then you have to go to E1, but if you came from right you have to go to E2. The version with two different routes shown on the lower part is fine and easy to learn.

The interpretation of the experiment was that memory representation in animals with a locomotory way of life is place-tied, rats being unable to learn different targets being on the same place if the place was reached by different routes. These behavioral results are to be explained according to Kardos by postulating a *mnemonic field* (Kardos 1988). It is like a photographic replica, and it contains in a way objects together with their locations, we should say today, in an egocentric frame of reference. (For a modern discussion of the frames of reference in human and animal locomotion, see Zaehle et al. 2007).

[2] Kardos was a strange kind of purist as language goes. On the one hand, he was fighting for the use of Hungarian language terminology in professional texts. At the same time, he was also creating his own Latinate or Greek-based terms for phenomena he thought related to his original contribution. Aequiterminal routes belong to this: its Latinate stem means 'equivalent ending'. The same innovation goes for his other key notion*, adiaphore determination,* as we shall see below.

Discrimination learning with a moving platform

Kardos went on to prove the importance of place learning in locomotory animals by using a discriminatory learning paradigm. In classical discriminatory learning situations when alternative doors are used, the position and the other cues like shape are randomized. Kardos, however, made position to be a crucial factor in itself. He used the experimental arrangement shown on Figure 4 (Kardos et al. 1978; Kardos 1988).

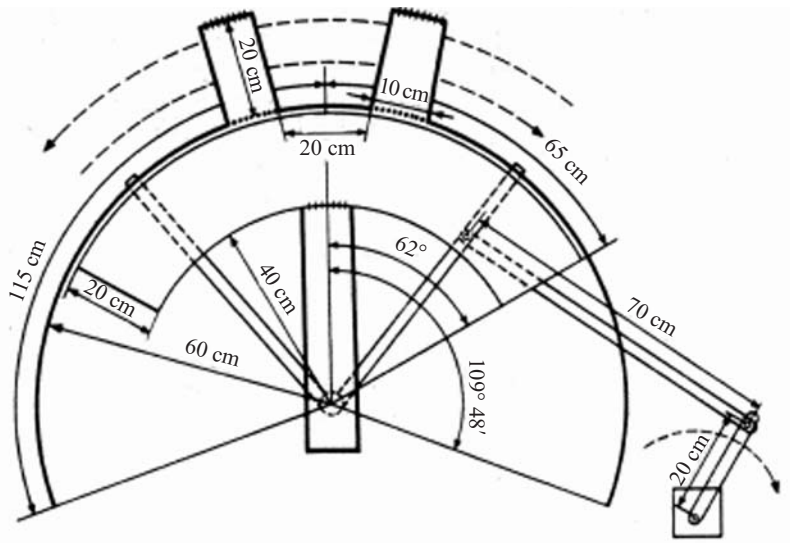


Figure 4. Discriminatory learning arrangement with a moving platform and moving doors (Kardos 1988)

Animals had to learn to chose the white door. The colors were alternating between left and right at each trial. The task was extremely difficult, required over 100 trials to learn. However, if the platform where the animal was standing, and the wall, where the doors were located both moved, the location cue disappeared and the task became rather easy. 17 trials were sufficient for the animals to learn the discrimination. In his own words:

> [t]o investigate whether discrimination learning is hindered by the emergence of a place-memory field, 10 Tolman rats were given learning tests in a maze in which a white door was a positive stimulus to be discriminated from a black door. Both doors were placed on a wall moving in respect to the environment. Results support the hypothesis that discrimination learning in situations with moving goal objects is significantly higher than such learning in situations with stationary goal objects. (Kardos et al. 1978: 101)

As Gilchrist (2010) puts it, "[a]s I understand it, Kardos is saying that when the platform is moving, the doors can no longer be identified by their spatial relationship to the larger room, because the doors are moving, and thus the animal is freed to pay attention to other aspects, like the color of the doors."

Using some more complex spatial learning situations such as star-shaped mazes, Kardos proposed a mnemonic theory slightly different from Tolman's (1948) cognitive maps. Kardos (1988) claimed that rodents basically maintain memory images as vivid as their percepts.

Kardos interpreted rodent memory in the framework of a locomotory ecology contrasted to the more manipulatory way of life of apes and humans. The advent of cognitive psychology was of course welcome by the senior cognitively oriented comparative psychologist.

His animal memory research had relatively little impact, though Jagaroo and Wilkinson (2008) refer to a technical aspect of the work of Kardos in their studies. Moving the stimulus situation was furthering discrimination learning.

The origins of mental life

In his late years, Kardos started an even more ambitious enterprise. Partly turning back to the Bühlerian heritage (Bühler 1927, 1934), he elaborated an even more general information- and guidance-related vision of the birth of the mind. In his theory, mental, or as he called them, neuropsychological phenomena arise in animal life when corollary, adjacent information about the environment becomes to be used and detached in a scheme he calls *adiaphore determination.*[3] Mental life gradually evolves as a consequence of using predictive information about noxious, harmful events available in stimulus arrays (Kardos 1980).

The Bühler heritage

This theoretical book of Kardos, when compared to all his other work, reads as surprisingly speculative. He is not concerned with experiments, and he does not quote many references. On the basis of some elementary biological background, Kardos sets out to analyze the postulated behavior of theoretical monocellular organisms. This excursion is used to shed light on the origin of mental life.

In this regard, it is remarkable that his teacher half a century earlier used the same attitude when proposing a unified sign-based framework for psychology. The unity of biological and meaningful elements in human life on all levels of mental organization was the key notion for Bühler:

> The distance between the integrated behavior of the amoeba and human scientific thought is certainly impossible to grasp. Still, on the basis of the most modern observations both can come under two common concepts: *they are holistically organized and are characterized by meaningful events.* (Bühler 1927: 392)

In his analysis of the origin of mind the starting point for Kardos is avoidance behavior. Warning signs are crucial in the development of mental life. Starting from the etymology of prevention ('prevent'® 'praevenio'), he claims that organisms use information that precedes

[3] Form the Greek word *adiaphoros,* meaning 'indifferent' (Kardos 1980: 19).

harmful events: “harmful impacts are consistently preceded by biologically irrelevant impacts” (Kardos 1980: 24).

Signals precede the harmful event. The animal avoids the harmful space, and „the adiaphore space is a secure starting place; from here, by well-controlled action, it can avoid any dangerous contact or can achieve contact when desirable” (Kardos 1980: 94).

Adiaphore determination in Kardos's work

Side effects precede harmful effects. These side effects become information, and form the basis of mental life. With all aspects of the environment, “some representative process is coordinated in the central nervous system in an approximately one-to-one relation” (Kardos 1980: 104). This information becomes functionally independent, and in this way „the mental phenomenon appears in animal life” (Kardos 1980: 104).

According to Kardos the mere causal signaling function becomes „pure” information. The key to understand mental life is to understand the genesis of information and its decontextualization, and to have an autonomous life.

The genesis of representation in modern cognitive theories

There are interesting parallels between Kardos's theory and some trends in modern cognitive theorizing (of which Kardos was not aware). Paul Dretske (1981, 1988) is his 1981 book *Knowledge and the Flow of Information* proposes a peculiar mental vs. non-mental distinction. Here, representations arise when certain neural constructions emulate the cause being still present.

The shift from indication – that is, the natural signifying relationship – to representation is similar to the one proposed by Kardos: the relative autonomy of the signifying relation and therefore the possibility of misrepresentation. For Kardos, this entails the possibility of error in mental representation.

In the philosophical analysis proposed, the representational functions would have a stepwise origin in phylogenesis: the creation of neural events that present a state of affairs to the organism as if it was true while it is not crucial in this chain of events. The causal event in this way becomes a carrier of information. Think of afterimages, for example.

There are of course many differences between the two views but the similarity in the genesis of information and the mind as a corollary event of reorganizing neural causal chains is a remarkable parallel.

Some coda

There are some aspects which make it relevant to remember Kardos's work.

1) *The Bühler heritage.* It is remarkable that a Central European tradition of the between-the-world-wars period can survive as a motivating metatheory in the context of dramatic social

changes. This metatheory tries to treat signaling in all mental phenomena as being crucial. This aspect shows up in all three fields of Kardos's work.

2) *A creative integration of the Pavlov tradition of 'signal systems' and learning theory.* All the social difficulties notwithstanding, Kardos managed to elaborate a rather original theory of types of learning and memory organization being related to the way of life of the different species.
3) *A mathematical information theoretic flavor* of treating issues in a substance neutral language.
4) A peculiar interpretation of the brain–mind relation: *neural becomes mental via becoming information.*

All of these were developments of a more or less self-imposed program, while Kardos tried to follow comparative psychological literature, though he had not had any particular knowledge about the parallel modern cognitive trends.

References

Barkóczi, I. (1998) Önarckép háttérrel. [Self-portrait with a background.] In: P. Bodor, Cs. Pléh, and G. Lányi (ed.) *Önarckép háttérrel. Magyar pszichológusok önéletrajzi írásai.* [Self-portrait with a Background. Autobiographies of Hungarian Psychologists.] Budapest: Pólya, 73–86.

Bodor, P., Pléh, Cs., and Lányi, G. (1998) *Önarckép háttérrel. Magyar pszichológusok önéletrajzi írásai.* [Self-portrait with a Background. Autobiographies of Hungarian Psychologists.] Budapest: Pólya.

Brunswik, E. (1934) *Wahrnehmung und Gegenstandswelt.* [Perception and the World of Objects.] Oxford, England: Deuticke.

Brunswik, E. (1956) *Perception and the Representative Design of Psychological Experiments.* 2nd ed. Berkeley: University of California Press.

Brunswik, E., and Kardos, L. (1929). Das Duplizitätsprinzip in der Theorie der Farbenwahrnehmung. *Zeitschrift für Psychologie* 111, 307–320.

Bühler, K. (1913) *Die Gestaltwahrnehmungen.* Stuttgart: W. Spemann.

Bühler, K. (1927) *Die Krise der Psychologie.* Jena: Fischer.

Bühler, K. (1934) *Sprachtheorie.* Jena: Fischer.

Crannell, C. W. (1948) Modification of the Kardos Shadow Experiment for Demonstrations of Color Mixing. *Science* 108 (2799), 190–191.

Dretske, F. (1981) *Knowledge and the Flow of Information.* Cambridge (MA): MIT Press.

Dretske, F. (1988) Explaining Behavior: Reasons in the World of Causes. Cambridge (MA): MIT Press.

Engels, F. (1876) *The Part Played by Labour in the Transition from Ape to Man.* Electronic source: http://www.marxists.org/archive/marx/works/1876/part-played-labour/index.htm.

Fleming, R. W., Torralba, A., and Adelson, E. H. (2004) Specular reflections and the perception of shape. *Journal of Vision* 4, 798–820.

Gilchrist, A. (2004) Lajos Kardos: Outstanding Hungarian Gestaltist. *Perception* 33. ECVP Abstract Supplement.

Gilchrist, A. (2006) *Seeing Black and White.* Oxford: Oxford University Press.

Gilchrist, A. (2010) Letter to Csaba Pléh. April 14th, 2010.

Gilchrist, A. L., and Annan, V. Jr. (2002) Articulation effects in lightness: Historical background and theoretical implications. *Perception* 31 (2), 141–150.

Gould, S. J., and Vrba, E. S. (1982) Exaptation: A missing term in the science of form. *Paleobiology* 8 (1), 4–15.

Hull, C. L. (1934). The concept of the habit-family hierarchy and maze learning. Parts I & II. *Psychological Review* 41, 33–54, 134–152.

Jagaroo, V., and Wilkinson, K. (2008) Further considerations of visual cognitive neuroscience in aided AAC: The potential role of motion perception systems in maximizing design display. *AAC: Augmentative and Alternative Communication* 24, 29–42.

Kardos, L. (1930) Diskussionen über Probleme des Farbensehens. Erwiderung an D. Katz. [Discussions of problems of color perception.] Reply to D. Katz. *Archiv für die Gesamte Psychologie* 78, 185–215.

Kardos, L. (1934) *Ding und Schatten.* Leipzig: Barth.

Kardos, L. (1935) Versuch einer mathematischen Analyse von Gesetzen des Farbensehens. Nähere Bestimmung des funktionalen Verhältnisses zwischen Farbenerlebnis und Reizgesamtheit. *Zeitschrift für Sinnesphysiologie* 66.

Kardos, L. (1959) Tanulás és emberréválás. [Learning and anthropogenesis.] *Pszichológiai Tanulmányok* 1, 41–56.

Kardos, L. (1960) Die Grundfragen der Psychologie und die Forschungen Pawlow's. Budapest: Akadémiai Kiadó.

Kardos, L. (1964) Kibernetika és pszichológia. [Cybernetics and psychology.] *Magyar Pszichológiai Szemle* 21, 523–529.

Kardos, L. (1965a) Az állatlélektani kutatások jelentősége és néhány elvi kérdése. [The importance and theoretical issues of zoopsychology.] *Pszichológiai Tanulmányok* VII, 105–113.

Kardos, L. (1965b) Az érzékletek átalakulásai. [Transformations of percepts.] *Pszichológiai Tanulmányok* VIII, 11–32.

Kardos, L. (1966) A korrekciós rendszerek szerepe az érzékleti szerveződésben. [The role of correction systems in perceptual organization.] *Pszichológiai Tanulmányok* IX, 11–20.

Kardos, L. (1980) The Origins of Neuropsychological Information. Budapest: Akadémiai Kiadó.

Kardos, L. (1984a) Errinerungen an Karl Bühler. In: A. Eschbach (ed.) *Bühler Studien.* Vol. 1. Frankfurt: Suhrkampf, 31–39.

Kardos L. (1984b) *Tárgy és árnyék.* [Object and shadow.] Budapest: Akadémiai Kiadó.

Kardos, L. (1988) *Az állati emlékezet.* [Animal memory.] Budapest. Akadémiai Kiadó.

Kardos L., and Barkóczi, I. (1953) "Aequiterminális" viselkedésrészletek jelentősége az állati tanulásban. [The importance of "aequiterminal" behavioral events in animal learning.] *MTA Biológiai Osztályának Közleményei* 2, 95–114.

Kardos, L., and De Renoche, I. (1966) Role of the Vestibular Apparatus in the Locomotor Functions of Rats. *Rivista di Psicologia* 60, 15–33.

Kardos, L., Barkóczi, I., and Kónya, A. (1971) An experiment in the direct comparison of learning locomotional and manipulational forms of action. *Magyar Pszichológiai Szemle* 28, 1–15.

Kardos, L., Da Pos, O., Dellantonio, A., and Saviolo, N. (1978) Discrimination learning and visual memory. *Italian Journal of Psychology* 5, 101–133.

Kohler, I. (1951) Über Aufbau und Wandlungen der Wahrnemungswelt. *Sitzungsberichte der österreichishen Akademie der Wissenschaften* 222.

Kohler, I. (1962) Experiments with goggles. *Scientific American* 206 (5), 62–86.

Kohler, I. (1963) The formation and transformation of the perceptual world. *Psychological Issues* 3 (4, Monograph No. 12), 1–173.

Köhler, W. (1921) Intelligenzprüfungen an Menschenaffen. Berlin: Springer.

Kovács, M. (1994) Liberal Professions and Illiberal Politics: Hungary from the Habsburgs to the Holocaust. Oxford: Oxford University Press.

Logvinenko, A. D., Petrini, K., and Maloney, L. T. (2008) A scaling analysis of the snake lightness illusion. *Perception & Psychophysics* 70, 828–840.

Murányi, G. (1985) "Amire talán a legbüszkébb vagyok". Múlt és jelenidézés Kardos Lajos professzorral. [What I am most proud of. Interview with Professor Lajos Kardos.] *Magyar Nemzet,* 19 January, 1985.

Pléh, Cs. (1985) Élmények, barátok, örömök: Interjú a 85 éves Kardos Lajossal. [Experiences, friends and pleasures: An interview with Lajos Kardos on his 85th birthday.] *Magyar Pszichológiai Szemle* 42, 345–351.

Pléh, Cs. (2008) *History and Theories of the Mind.* Budapest: Akadémiai Kiadó.

Tánczos, Zs. (1977) Kardos Lajos munkássága a színkonstancia területén. Eredményeinek jelentősége a pszichológiai elméletalkotásban. [The work of Lajos Kardos in the domain of color constancy. The relevance of his results is psychological theorizing.] *Magyar Pszichológiai Szemle* 34, 517–535.

Todorovic, D. (2010) Personal correspondence with Csaba Pléh.

Tolman, E. C. (1932) *Purposive Behavior in Animals and Men.* New York: Century.

Tolman, E. C. (1948) Cognitive maps in rats and men. *Psychological Review* 55, 189–208.

Zaehle, T., Jordan, K., Wüstenberg, T., Baudewig, J., Dechent, P., and Mast, F. W. (2007) The neural basis of the egocentric and allocentric spatial frame of reference. *Brain Research* 1137, 92–103.

HISTORICAL PERSPECTIVES ON THE *WHAT* AND *WHERE* OF COGNITION

Lena Kästner and Sven Walter

What is cognition? The embarrassing answer is: There is no unanimously accepted answer, not even remotely. We simply don't seem to know.

We don't know *yet,* optimists insist. We have a list of fairly uncontroversial prototypes of cognitive processes – categorization, learning, perception, reasoning, etc. – and it is only a matter of time until we can specify a "mark of the cognitive" capturing their common core. The problem is: There is no reason for thinking the mechanisms implementing these prototypes will have anything significant in common, materially or functionally.

We will *never* know, pessimists insist. "Cognition" is merely a label for a motley bundle of processes that are of interest to cognitive scientists for some reason or other. Many sciences invoke key concepts that lack crisp and clear definitions – "language" in linguistics, say, "gene" in biology, or "intelligence" in psychology. "Cognition" may just be another case where researchers recognize a phenomenon if they come across it, but are unable to provide necessary and sufficient conditions.

But this agnosticism cannot succeed. We are interested not only in the *what,* but also in the *where* of cognition. Clark and Chalmers (1998) famously maintained that the material vehicles of some cognitive processes *extend* beyond the brain and the body into the environment. Defenders of a more conservative view, in contrast, insist that cognition resides in brains, or at least bodies. The only sensible way of settling this dispute is to provide a mark of the cognitive, and then go and see where in the world the processes fulfilling it are found. Resolving the *where*-question thus presupposes resolving the *what*-question. This is why agnosticism is untenable – we must know *what* cognition is before we can make any progress on its *where.*

Agnosticists may try to dismiss the *where*-question, arguing that extended cognition is merely a fancy philosophical hypothesis. A look at the history of cognitive science shows that this is not true. The idea of cognitive extension is a natural consequence of earlier approaches to cognition, and it is a legitimate question to ask whether it is correct or not. A look at the history also allows us to understand the importance of, and the various answers to, the *what*- and *where*-questions (for a more detailed version of the following see Walter and Kästner 2012).

Classicism

Scientific work on computation, information theory, and cybernetics during and after World War II culminated in the interest in the artificial design of intelligent agents that gave AI

its name. Two "classicist" views of cognition originated from this early work: the more theoretically inspired algorithmic "rules and representations" approach of *good old-fashioned artificial intelligence* (GOFAI), and the more biologically inspired neural network approach of *connectionism.*

GOFAI's answer to the *what*-question was that cognition is algorithmic *information processing* in the sense of rule-governed, sequential *computations* over structured symbolic *representations.* Since in humans these representations are arguably encoded neurally, the computational processes in question are an entirely intracranial affair. GOFAI's answer to the *where*-question thus was: in the head. The world serves only as a source of perceptual input, and the arena for behavioral output, while all the cognitive processing is done in the head; cognition is a central element "sandwiched" (Hurley 1998) between the peripheral buffer zones of perception and action.

According to connectionism, cognition is grounded in spreading activation in heavily connected networks of neuron-like information processing units. *Information processing* thus again played a crucial role, and so did *computation.* Computations in connectionist networks, however, are not rule-based (not explicitly, at least) they are *local,* i.e., they take place at the level of individual network nodes, and thus *parallel* in the sense that multiple nodes are simultaneously active. A single node is usually involved in a range of a network's states, and its current activity is determined by the network's overall activation pattern. Although this starkly contrasts with GOFAI's conception of computation as a rule-based, global, sequential process, it is nevertheless computation. *Representations* also remained a crucial element. But since individual nodes do not normally map in a one-to-one fashion onto the constituents of what the network's overall state stands for, they were said to be *subsymbolic* or *distributed* representations. Hence, although the details were different, connectionism's answer to the *what*-question was essentially the one already given by GOFAI: information processing by computations over representations. Connectionism obviously also endorsed the sandwich model. Moreover, since the relevant networks in humans are their brains, the answer to the *where*-question was the same, too: cognition is an entirely intracranial affair.

Dynamicism

Classicism was most successful in modeling disembodied, abstract features of human cognition that can be performed "off-line," i.e., detached from the world, like inference drawing, and problem-solving (GOFAI), or pattern recognition (connectionism). In contrast, advocates of a dynamicist approach emphasized the importance of "on-line" cognition: cases where cognitive systems are dynamically coupled to their environments in immediate, real-time interactions, and under continuous reciprocal causal influence (the distinction between "off-line" and "on-line" cognition is in Wheeler and Clark's (1999) sense). Dynamicists stressed that brains are seamlessly integrated into their bodily and extrabodily environments in such a way that neurophysiological, physiological, and environmental processes form a single, dynamically changing whole. We should therefore treat cognitive systems as *dynamical systems,* and model them by sets of differential equations: "cognitive agents are dynamical systems and can be scientifically understood as such" (van Gelder 1999: 13). As a consequence, dynamicists downplayed the role of *computation:* "[r]ather than computers,

cognitive systems may be dynamical systems; rather than computation, cognitive processes may be state-space evolution within these very different kinds of systems" (van Gelder 1995: 346). They also eschewed the appeal to *representations:* since dynamical systems are continuously evolving, there are no discrete, sequential steps in which one representation is transformed into another (van Gelder 1995): "[w]e are not building representations of the world by connecting temporally contingent ideas. We are not building representations at all! Mind is activity in time…" (Thelen and Smith 1994: 338). Yet, mathematically speaking, dynamical systems are characterized by sets of state variables and sets of laws determining how the values of these variables change over time. Each possible state of a system is a point in its state space, and a sequence of states is a trajectory through that state space. One may thus argue, for instance, that these trajectories through state space are, albeit in a weak sense, representations of the system's behavior.

The dynamicist's official view on the *what*-question is: cognition is state-space evolution in a dynamical system, and thus it is neither decidedly computational nor decidedly representational, although still fundamentally information processing. The sandwich model is rejected, for the world is more than merely the passive source of input for, and receiver of output from, the cognitive system: "[t]he cognitive system does not interact with other aspects of the world by passing messages or commands; rather, it continuously coevolves with them" (van Gelder and Port 1995a: 2). Dynamicists may also be read as offering a radical answer to the *where*-question. Since cognitive systems are a dynamically changing whole comprising brain, body, and environment, the skull no longer looks like a natural boundary for the cognitive: "[c]ognitive processes span the brain, the body, and the environment" (van Gelder and Port 1995b, ix). Understood that way, dynamicism seems to support the idea of *extended cognition.* However, as dynamical systems are abundant, cognitive processes can at best be a *subset* of dynamical processes; and unless dynamicists specify what exactly it is that makes a dynamical process cognitive, their answer to the *where*-question may be interpreted conservatively: although dynamical processes are everywhere and crisscrossing the boundaries of brain, body, and environment, the dynamical processes that are cognitive may all reside in the brain.

Situated cognition

Dynamicists criticized the insular, sandwiched view of cognition characteristic of classicism, and instead stressed the importance of *body* and *environment.* Like dynamicism, situated approaches to cognition argued that classicism focused too narrowly on abstract programs for specialized feats of reasoning and inference in highly specialized domains, thereby neglecting that cognition emerges "on-line" out of the interactions between embodied cognitive systems and their environments, rather than being done "off-line" by a detached computational and representational system implemented in the brain. Understanding cognition thus requires understanding how physically embodied agents achieve sensorimotor control in "fluid and flexible" (Wheeler 2005: 170) real-time interactions with their environment. The slogan was to put cognition "back in the brain, the brain back in the body, and the body back in the world" (Wheeler 2005: 11). The resulting situated approaches to cognition are a relatively recent development with a variety of subtly different strands whose key tenets, theoretical

and terminological commitments, and interrelationships, and interdependencies are still in disarray (Robbins and Aydede 2009). Below we offer a taxonomy that strikes us as plausible (Walter 2010a), but others may disagree about the correct classification of the various approaches.

Embodied cognition

According to the embodied approach, cognition bears a profound relation to bodily processes in the sense that "the specific details of human embodiment make a special and […] ineliminable contribution to our mental states and properties" (Clark 2008b: 39). Pioneering work in the embodiment paradigm came from Rodney Brooks' bottom-up robotics (Brooks 1991), but quite generally the embodied approach to cognition is the attempt to carry out what Anderson (2003) called the "physical grounding project," viz. to show exactly how an agent's physical features and abilities contribute to his or her cognitive processing. Within the embodied cognition paradigm, the grounding relation is spelled out in at least two different ways.

Paradigm examples of the embodied approach to cognition are Lakoff and Johnson's (1999) work on our body's contribution to our conceptual repertoire, and McBeath et al.'s (1995) study on how baseball outfielders manage to catch fly balls. According to Lakoff and Johnson (1999), all our concepts are ultimately derived from basic concepts that stem directly from, and are constrained by, the type of body we possess (e.g., spatial ones like *up, down, front, back, etc.*). According to McBeath et al. (1995), a classicist sandwich solution to the problem of catching a fly ball would be to take the visual perception of the ball as input, generate an internal representation, let an internal reasoning system use that representation to compute the ball's future trajectory, and finally trigger an appropriate motor output. In reality, McBeath et al. argue, the solution relies on certain characteristics of the outfielder's body, thereby minimizing the need for internal computation and representation: simply run in such a way that the optical image of the ball appears to present a straight-line constant speed trajectory against the visual background.

Both examples illustrate how "the presence of a humanlike mind *depends* quite directly upon the possession of a humanlike body" (Clark 2008b: 43; emphasis added). The embodied approach therefore rejects the sandwich model. A straightforward positive answer to the *what*-question, however, is lacking. On the one hand, as the research of McBeath et al. illustrates, there is a certain depreciation of the role of computation and representation (also highlighted in Brooks' work on bottom-up architectures in robotics), but, on the other hand, Lakoff and Johnson's work is rather neutral with regard to these issues. No clear answer to the *what*-question, thus. The answer to the *where*-question, in contrast, is clear, and it is still conservative: Cognitive processes, although *causally dependent* upon extracranial bodily processes, are an entirely intracranial affair.

According to a stronger version of the embodied approach, cognitive processes are not only *dependent* upon but actually *constituted* by bodily processes. Support for this stronger claim comes from studies showing that vision essentially relies on bodily movements (e.g., Ballard et al. 1997; Noë 2004; O'Regan 1992; O'Regan and Noë 2001). Shapiro argues that bodily movements are not only extracranial aids but "as much *part* of vision as the detection

of disparity or the calculation of shape from shading" (2004: 188), so that "[v]ision for human beings is a process that *includes* features of the human body" (2004: 190; both emphases added). As in the case of the weak embodied approach, the implications with regard to the role of computation and representation are unclear. While Noë (2004) is usually pictured as a strict anti-representationalist, Ballard et al. (1997: 735) are clearly in favor of representations and argue that their model "strongly suggests a functional view of visual computation" able to combine the idea that vision is a form of acting with the idea that it is a computational process. The strong embodied approach gives a more radical answer to the *where*-question, however: Cognitive processes include *extracranial bodily* processes and are thus not merely in the head. This ambivalence in the embodied approach is rarely noted in the literature, although it has obvious ramifications for the proper study of cognitive processing. According to the weaker version, cognitive processes are restricted to an organism's brain, while according to the stronger, they are leaking out into the organism's body.

Embedded cognition

Embedded approaches to cognition stress the role of the environment and its active structuring by the agent. Recent research on visual processing (Noë 2004; O'Regan, and Noë 2001) suggests that instead of creating detailed internal representations as the basis for later-stage cognitive processing, human subjects extract the relevant information "on the fly" from the world itself. Kirsh and Maglio's (1994) research on "epistemic actions" highlights a similar kind of environmental "offloading" or "outsourcing" of cognitive load. Experienced *Tetris* players rotate the figures on the screen rather than mentally because it is cognitively less demanding. The important point is not just that the way the world is influences an agent's cognitive processing; it is that the agent himself or herself *actively* structures his or her environment in order to facilitate cognitive processing. The embedded approach may presuppose the embodied approach in the sense that an agent's capacity for cognitive offloading depends not only on the environment but also on his or her body because it is his or her body which determines how he or she can perceive, navigate, and manipulate her surroundings.

Regarding a positive answer to the *what*-question, the embedded approach is – again – not particularly forthcoming. What is clear is that while the *computational* nature of cognition is typically not denied, the idea of off-loading shows that the computations in question may directly involve extrabodily items rather than their rich internal *representations:* Why bother representing something internally that is right there in your environment? Simply use, as Brooks famously put it, the world itself as "its own best model" (Brooks 1991: 583). The embedded approach is decidedly more radical than the two embodied approaches, for it entails that cognitive processes must be studied not by looking at their causal (weak) or constitutive (strong) grounding in extracranial bodily processes, but by looking at the way an agent uses his or her environment's structure, or actively structures his or her environment. Regarding the *where*-question, the embedded approach is a natural extension of the weak embodied approach. Like the weak embodied approach, it specifies the grounding relation in terms of *causal dependence;* unlike the weak embodied approach, however, it takes the dependence base of cognitive processes to contain not only extra*cranial bodily,* but also extra*bodily* pro-

cesses. Cognitive processing thus takes place in the brain and in the extracranial parts of the body, although it causally depends upon the extrabodily environment.

Extended cognition

From embodied and embedded cognition, it is only a short step to extended cognition. If body and environment are indeed crucial for cognitive processing, then they may literally be a *part* of – rather than merely causally contributing to – cognition. Just as the embedded approach extends the dependence base of the weak embodied approach from extracranial bodily processes to extrabodily processes, extended cognition is a natural corollary of the strong embodied approach. Like the strong embodied approach, it stresses that cognitive processes are partially *constituted* by extracranial processes; unlike the strong embodied approach, however, the constituents of cognitive processes are taken to be not only extra*cranial,* but also extra*bodily.*

Regarding the *what*-question, most advocates of the extended approach would apparently be prepared to endorse the claim that cognition is a computational information processing process, albeit one which consists in computations over internal or external representations, or even the extrabodily items themselves: "at least some of the computational systems that drive cognition reach beyond the limits of the organismic boundary" (Wilson 2004: 165). The extended approach is thus surprisingly conservative. All it does is to allow for computational processes to range not only over internal representations, as in classicism, but also over external representations, or the extrabodily items themselves. Unlike classicism, however, the extended approach rejects the sandwich model, for the world is an active part of cognition, not only the passive source of input and the stage for output. With regard to the *where*-question, the extended approach is the most liberal: Cognitive processing involves intracranial, and extracranial bodily and extrabodily processes.

The what of cognition, again

This admittedly brief overview shows that the extended approach is not merely a fancy philosophical idea with no basis in cognitive scientific practice. There are two routes to the idea of cognitive extension, one *via* dynamicism, and one *via* embodied and embedded approaches. The *where* of cognition is thus a substantial and important issue that needs to be resolved. As said in the beginning, agnosticism is not a viable option because answering the *where*-question arguably requires answering the *what*-question (Walter 2010b). Unfortunately, as the preceding considerations have shown, there is not even a remotely unanimously accepted answer to the question *"What is cognition?,"* except for the idea that cognition probably has something to do with information processing, which can at best be a necessary, not a sufficient condition. Table 1 summarizes the results.

Table 1. TITLE

	What	Where	Sandwich	Information Processing	Computation	Representation
GOFAI	algorithmic information-processing, rule-based, sequential computations over symbolic representations	intracranial	yes	yes	yes global sequential rule-based	yes structured symbolic
Connectionism	spreading neural network activation	intracranial	yes	yes	Yes local parallel not (explicitly) rule-based	yes subsymbolic distributed
Dynamicism	state-space evolution in dynamical systems	*radical interpretation*: intracranial plus extracranial bodily plus extrabodily *conservative interpretation*: intracranial	no	yes	no	officially no maybe in a weaker sense
Weak Embodied Cognition	*no explicit answer*	intracranial	no	arguably yes	*some studies*: no need for (much) computation *some studies*: neutral	*some studies*: no need for internal representations *some studies*: neutral, maybe grounded in bodily features
Strong Embodied Cognition	*no explicit answer*	intracranial plus extracranial bodily	no	arguably yes	*some studies*: no need for (much) computation *some studies*: neutral	*some studies*: no need for internal representations *some studies*: neutral
Embedded Cognition	*no explicit answer*	intracranial plus extracranial bodily	no	arguably yes	not always required Ballard *et al.*: yes	Noë: no *others*: neutral or yes
Extended Cognition	information processing involving computations over internal representations, external representations or extrabodily items	intracranial plus extracranial bodily plus extrabodily	no	yes	yes	not necessarily

Thus far, it seems, we do not know what is distinctive about cognition, nor do we understand how it works, or where to look for it. The situation seems rather bleak. What are the options? Rather than immediately delving into the details of specific accounts of cognition, we suggest that it may be worthwhile to first ask a more general question: Should "cognition" be taken to be a *natural kind term,* a *cluster term,* or an *umbrella term?*

"Cognition" as a natural kind term

Perhaps the most intuitive approach to answering the *what*-question is to assume that "cognition" is a natural kind term whose instances have a scientifically discoverable essence – what Adams and Aizawa (2008; 2009) famously call a "mark of the cognitive." Cognitive processes, they argue, "are natural kinds of processes" (2008: 80) consisting in computational operations that involve *non-derived representations,* and are implemented by special kinds of *mechanisms.* Since non-derived representations and the kinds of mechanisms at issue are found, currently at least, only in the brain, there is "defeasible reason to suppose that cognitive processes are typically brain bound and do not extend from the nervous system into the body and the environment" (2008: 70). Adams and Aizawa's natural kind conception of the cognitive thus entails a conservative answer to the *what*-question: Cognitive processes are, as a contingent matter of fact, found in the head, and only in the head.

Although this is not the place to go into the details, note that a number of complaints can be leveled against Adams and Aizawa's approach. First, since there is no received theory of non-derived content, we cannot tell whether a process fulfills their mark or not. Second, unless there is a theory of non-derived content, it is hard to substantiate their claim that non-derived representations are currently found in the brain alone and not in the brain *cum* body *cum* environment. Third, Adams and Aizawa's claim that cognitive mechanisms must be individuated in terms of their material implementation begs the question against the functionalistic approach usually adopted by defenders of an extended approach (Walter 2010b). Finally, prematurely equating cognition with specific kinds of brain processes forecloses fruitful future discoveries in cognitive science.

Despite this skepticism about the approach of Adams and Aizawa, they do seem to have a point. Any conception of the cognitive that could support an extended view would arguably have to cover so heterogeneous processes that "cognition" would fail to pick out a natural kind. In other words, if "cognition" is a natural kind term, the answer to the *where*-question is most likely going to be conservative.

"Cognition" as a cluster term

Given the broad range of phenomena we usually count as cognitive, and given our apparent difficulties in capturing a common essence, "cognition" may simply fail to pick out a natural kind. There may just not be a set of individually necessary and jointly sufficient conditions for a process to be cognitive. Since "cognition" could be a *cluster term,* any cognitive process could still share some of its characteristics with other cognitive processes, but they would only amount to a *family resemblance.* Wheeler (2005), for instance, suggests that cognitive

processes can be implemented by (1) non-computational and non-representational, (2) non-computational and representational, and (3) computational and representational mechanisms. In that case, finding a "mark of the cognitive" would mean identifying the structure of the underlying resemblances rather than a single common core that fixes the meaning of the term "cognition." Such an approach would seem to be compatible with all sorts of answers to the *where*-question, depending upon where in the world the processes in question are found.

"Cognition" as an umbrella term

A third way of thinking about the demarcation of the cognitive would be to loosely characterize the basic commonalities of individual cognitive phenomena while accepting that they neither form an overarching natural kind of "the cognitive", nor exhibit any family resemblances. "Cognition" would then be an *umbrella term* under which, as Clark proposes, a "motley crew of mechanisms" (Clark 2008a) finds shelter. The concept of "memory" may illustrate the idea. It was only after discovering dissociations in neuropsychological patients like H. M. that the different mechanisms underlying what formerly seemed to pick out a single general capacity for memory had been recognized. Subsequent research revealed that memory decomposes into a variety of phenomena, the most general distinction being between *long-term memory* (LTM) and *short-term memory* (STM), both of which allow for further subdivisions. Analogously, cognition may decompose into a range of causally distinct processes "with not even a family resemblance" (Clark 2008a: 95).

Decomposing cognition into subtypes may seem fairly reasonable once we reconsider the broad range of phenomena cognitive scientists are interested in. The important open question, however, is what unifies the umbrella's subtypes if there are not even family resemblances. Clark (2008a) suggests that cognition should best be understood as *information processing tightly coupled* to a *cognitive core* – probably, but not necessarily, the brain. However, Clark's characterization of the cognitive requires an account of "information," "coupling," and "cognitive core," and it is hard to see how to spell out the notion of a "*cognitive* core" without having already at hand a notion of the cognitive. Most importantly, however, Clark's approach is an instance of what we have dubbed "agnosticism." And as we have indicated in the beginning, it is a mistake to think that the *where*-question can be successfully answered against the background of an agnostic stance towards the *what*-question.

Conclusion

We started with the claim that there is at least one good reason to ask the *what*-question: In order to settle the *where*-question*,* we *have to* answer the *what*-question. Once we have done that, we can simply go and look where in the world we find cognition.

This means we need a mark of the cognitive. The prospects for such a mark, however, depend on what sort of term "cognition" is, as Section 4 has shown. Moreover, since each of the paradigmatic options presented here struggles with shortcomings, the situation seems rather bleak. Being aware of the options available to us, however, may help to guide cognitive scientific research and eventually enable us to answer both the *what*- and the *where*-questions.

References

Adams, F., and Aizawa, K. (2008) *The Bounds of Cognition.* Oxford: Blackwell.

Adams, F., and Aizawa, K. (2009) Why the mind is still in the head. In: P. Robbins, and M. Aydede (eds.) *The Cambridge Handbook of Situated Cognition.* Cambridge: Cambridge University Press, 78–95.

Anderson, M. (2003) Embodied cognition: A field guide. *Artificial Intelligence* 149, 91–130.

Ballard, D., Hayhoe, M., Pook, P., and Rao, R. (1997) Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences* 20, 723–767.

Brooks, R. (1991) Intelligence without reason. Proceedings of the 12th International Joint Conference on Artificial Intelligence, 569–595.

Clark, A. (2008a) *Supersizing the Mind.* New York: Oxford University Press.

Clark, A. (2008b) Pressing the flesh: A tension in the study of the embodied, embedded mind? *Philosophy and Phenomenological Research* 76, 37–59.

Clark, A., and Chalmers, D. (1998) The extended mind. *Analysis* 58, 7–19.

van Gelder, T. (1995) What might cognition be if not computation? *Journal of Philosophy* 92, 345–381.

van Gelder, T. (1999) Defending the dynamical hypothesis. In: W. Tschacher, and J. Dauwalder (eds.) *Dynamics, Synergetics, Autonomous Agents.* Singapore: World Scientific, 13–28.

van Gelder, T., and Port, R. (1995a) It's about time: A perspective to dynamical system approach to cognition. In: R. Port, and T. Van Gelder (eds.) *Mind as Motion.* Cambridge (MA): MIT Press, 1–43.

van Gelder, T., and Port, R. (1995b) Preface. In: R. Port, and T. Van Gelder (eds.) *Mind as Motion.* Cambridge (MA): MIT Press, vii–x.

Hurley, S. (1998) *Consciousness in Action.* Cambridge (MA): Harvard University Press.

Kirsh, D., and Maglio, P. (1994) On distinguishing epistemic from pragmatic action. *Cognitive Science* 18, 513–549.

Lakoff, G., and Johnson, M. (1999) *Philosophy in the Flesh.* New York: Basic Books.

McBeath, M., and Shaffer, D., and Kaiser, M. (1995) How baseball outfielders determine where to run to catch fly balls. *Science* 268, 569–573.

Noë, A. (2004) *Action in Perception.* Cambridge (MA): MIT Press.

O'Regan, K. (1992) Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology* 46, 461–488.

O'Regan, K., and Noë, A. (2001) A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences* 24, 939–960.

Robbins, P., and Aydede, M. (2009) A short primer on situated cognition. In: P. Robbins, and M. Aydede (eds.) *The Cambridge Handbook of Situated Cognition.* Cambridge: Cambridge University Press, 3–11.

Shapiro, L. (2004) *The Mind Incarnate.* Cambridge (MA): MIT Press.

Thelen, E., and Smith, L. (1994) A Dynamic Systems Approach to the Development of Cognition and Action. Cambridge (MA): MIT Press.

Walter, S. (2010a) Locked-in syndrome, BCI, and a confusion about embodied, embedded, extended, and enacted cognition. *Neuroethics* 3 (1) 61–72.

Walter, S. (2010b) Cognitive extension: The parity argument, functionalism, and the mark of the cognitive. *Synthese* 177 (2), 285–300.

Walter, S., and Kästner, L. (2012) The where and what of cognition: The untenability of cognitive agnosticism and the limits of the motley crew argument. *Cognitive Systems Research* 13 (1), 12–23.

Wheeler, M. (2005) *Reconstructing the Cognitive World.* Cambridge (MA): MIT Press.

Wheeler, M. (manuscript) *Extended X.* http://www.philosophy.stir.ac.uk/staff/m-wheeler/ExtendedX.php. Accessed October 20, 2009.

Wheeler, M, and Clark, A. (1999) Genic representation: Reconciling content and causal complexity. *British Journal for the Philosophy of Science* 50, 103–135.

Wilson, R. (2004) *Boundaries of the Mind.* Cambridge: Cambridge University Press.

ERNST MACH AND GEORGE SARTON:[1] HISTORY OF SCIENCE AS METAPSYCHICAL METHOD[2]

Hayo Siemsen

> *But in my opinion […] there are still deeper reasons why the scientist should give his attention to the history of science. I am thinking of those, which have been so splendidly illustrated by Ernst Mach in his* Mechanics*. (George Sarton)*

George Sarton had a strong influence on the modern history of science. The method he pursued throughout his life was the method he had discovered in Mach's *Mechanics* when he was a student in Ghent. This influenced him so much that he wrote his dissertation on the same topic, mechanics. In 1911, he even wrote a letter to Mach, asking him to join the editorial Board of ISIS, which Mach had to refuse at the time because of his health. Nevertheless, with Wilhelm Ostwald and Jacques Loeb, others relatively close to Machian thinking had an important influence on ISIS in the beginning.

As Marc de Mey has indicated in his introductory speech at the George Sarton Centennial in Ghent 1984, throughout his life Sarton was in fact implementing a research program inspired by the epistemology of Mach. Sarton in turn inspired many others (James Conant, Thomas Kuhn, Gerald Holton, etc.). What were the origins of these ideas in Mach and what can this origin tell us about the history of science and technology nowadays?

The following article will elaborate the epistemological questions which Darwin's "Origin" raised concerning human knowledge and scientific knowledge, and which Mach was first to answer. Sarton had an unusual concept of "genesis and development", which he proposed as the major goal of ISIS in his first editorial. Mach had elaborated this epistemology in his *Knowledge and Error,* which Sarton read in 1911 (de Mey 1984).

According to this epistemology, history becomes not only a subject of science, but a method of research and epistemology. Culture and science as part of culture are a result of a genetic process. In a double dependency, history of science shapes science, but its epistemology needs to be adapted to scientific facts and the philosophy of science. Sarton was well aware of the need to develop the history of science and the philosophy of science along the lines of this double dependency. Looking again at the origins of the central questions in the thinking of Ernst Mach, which gave rise to this research program, will help in the current epistemic

[1] A different version of this article with the same idea, but a focus on education instead of psychology has been published in Siemsen, H. (2012a) Ernst Mach, George Sarton and the Empiry of Teaching Science Part I. *Science & Education*, 21/4: 447-484 and Siemsen, H. (2012b) Ernst Mach, George Sarton and the Empiry of Teaching Science Part II. *Science & Education*, online first 25/04/2012.

[2] *Acknowledgements:* I would like to thank especially Marc de Mey, Gerald Holton, Richard Kremer and Michal Kokowski for the information they provided on specific historical details discussed in this article. The idea for the article and its intellectual setting came from discussions especially with Csaba Pléh, Lilia Gurova, Nicole Rossmanith, and Andreas Reichelt. I would also like to thank my father, Karl Hayo Siemsen for his invaluable help in proofreading.

and methodological development of the history of science and technology, especially because these origins are seemingly mostly forgotten today.

George Sarton, Ernst Mach, and the intuitive idea of the *Mechanics*

In 1883, Ernst Mach wrote a book entitled *The Science of Mechanics – A Critical and Historical Account of Its Development*.[3] It was to become a very influential book. It led to fundamental changes in the area of physics by questioning the basic assumptions of the then-dominant Newtonian mechanics. This enabled the development of the "new physics", i.e., quantum physics and relativity theory. The *Mechanics* was also influential as a method of science. Mach's historio-critical or historio-genetic method of writing the book was upheld by many as *the* method of doing science (for instance, Einstein 1916). Furthermore, it inspired theoretical approaches to science, such as the one from George Sarton.

Sarton had been influenced by Mach's ideas very early in his career. He had already read Mach's *Mechanics* during his studies at the University of Ghent. For instance in 1908–1910, he participated at Paul Mansion's courses on "Histoire des sciences mathématiques et physiques", which were especially focused on the history of geometry, on Poincaré and Mach, especially Mach's *Space and Geometry.* Sarton wrote his dissertation in 1911 on "Les Principes de la Méchanique de Newton" (which unfortunately is lost, but given the closeness of the topic and Sarton's (and his teacher's) interest in Mach, it was very probably related to Mach's *Mechanics.* In 1913, Sarton even wrote a letter to Mach, asking him to participate in the founding board of ISIS. Mach probably declined, as his health was already far too fragile at the time. But with names such as Henri Poincaré, Wilhelm Ostwald, Svante Arrhenius or Jacques Loeb, the first editorial board of ISIS reads like a "who is who" of people who had been strongly influenced by Mach's epistemology. Maybe Mach in his (albeit lost) return letter had actually suggested some of the names to Sarton.[4] Directly after his emigration to the US, Sarton often used Mach in quotations, especially regarding scientific methodology and his fundamentally important idea of the teaching of the history of science as a means of science education.[5] It was specifically these ideas, which strongly influenced Sarton's intellectual successors, such as Conant, Kuhn, etc.

In his introductory article of ISIS in 1911, Sarton quotes from Mach's "Knowledge & Error"[6] regarding the meaning of ISIS' central goal of *genesis and development:* "Mach stated very correctly that 'the formation of scientific hypotheses is merely a further degree of development of instinctive and primitive thought, and all the transitions between them can be

[3] The German title is *Die Mechanik – historisch-kritisch dargestellt*. For a later English edition (1893), Mach asked the translator McCormack to change the English subtitle into "genetic" instead of "critical and historical" (see letters from Mach to McCormack at the Paul Carus archive, Bloomingdale). For printing reasons, this idea eventually had to be dropped.

[4] Mach tended to be very courteous to everybody. Even at times when he was very ill, he still tried to answer his letters, also those from unknown young scientists.

[5] According to Marc de Mey, this influence is supported by letters in the Sarton archive, which still needs to be researched in detail.

[6] This was Mach's most epistemological book, subtitled: *Sketches to the Psychology of Research.*

demonstrated.'" Later, Sarton (1916, 1918) specifically described Mach's genetic method as the centre of his own historical method:

> The purpose of the history of science is to establish the *genesis* and the *development* of scientific facts and ideas, taking into account all intellectual exchanges and all influences brought into play by the very progress of civilization. It is indeed a history of civilization from its highest point of view. The center of interest is the *evolution* of science, but general history remains always in the background. But in my opinion […] there are still *deeper reasons* why the scientist should give his attention to the history of science. I am thinking of those which have been so splendidly illustrated by *Ernst Mach in his Mechanics.* For one thing it is obvious that they that know the entire course of the development of science will, as a matter of course, judge more freely and more correctly of the significance of any present scientific movement than they who, limited in their views to the age in which their own lives have been spent, contemplate merely the momentary trend that the course of intellectual events takes at the present moment. […] Moreover […] this *critical* work is essentially of an historical nature. While it helps to make the whole fabric of science more coherent and more rigorous, at the same time it brings to light all the accidental and conventional parts of it, and so opens new horizons to the discoverer's mind. […] The few serious courses that have been thus far devoted to these studies, here and abroad, have been, with the possible exception of *Mach's lectures,* far too philosophical […]. As a matter of fact, no history of science has ever been written from this point of view – none that I know of, not even *Ernst Mach's admirable history of mechanics,* although he has come considerably nearer to this ideal than any other author.

Sarton here clearly sees his own endeavors fundamentally as a continuation of Mach's genetic idea of teaching.

The psychology of Mach's Mechanics

Sarton's and Mach's ideas are thus very similar in many aspects, but there are – as Sarton had explicitly noted in the quotation before – also differences. What are the main differences between Mach's and Sarton's ideas regarding the history of science? Both achieved a major impact with their works. Sarton certainly had a huge influence on the history of science itself as a discipline, and the methodology of that discipline. He arguably also had some effects on science teaching, especially via James Conant and his General Education movement, even though the latter effects seem to have been limited in time (see Holton 2003). Mach had many influences in different scientific areas in addition to physics and he had a strong long-term influence on education at least in some countries in the world (e.g., Finland, see Siemsen and Siemsen 2009). Therefore, there seem to be two areas in which the difference of ideas shows "empirically": teaching science and the general method of scientific thinking (i.e., *Erkenntnistheorie*)[7] with its different applications.

[7] The German *Erkenntnistheorie* literally translates as "theory of knowledge/cognition". The meaning in the Machian sense is different from but related to epistemology (which will be elaborated later). Because of this

What could have been the reason for these differences? Sarton gives a hint on this in his analysis of what he supposes as Mach's later direction of developing his method: "[…] I do not know of any course in which such demonstrations have been actually carried out. The reader will surely think of Ernst Mach […]. I have no definite information about his method of teaching; I do not know to what extent his courses were *experimental.* But as Mach had become more and more interested in *psychological* rather than historical research, it is likely that his teaching was very different from the one of which I am thinking."

From a psychology of science view, one can even observe a slight desperateness in Sarton's regret regarding his ignorance of Mach's method of teaching. This missing information might have been at the basis of the differences between his and Mach's ideas. It could well be at the basis of some of the empirical "difficulties" of Sarton's ideas. Two categorical differences seem to be central: One is Sarton's opposing of psychological versus historical research (instead of using a Machian monistic, general approach, integrating both). The other is the importance attributed by Sarton to experiments. As we shall see, the two are actually closely interrelated, experimentalism[8] being a result of specific psychological assumptions.

Why did Sarton see this special role for experiments? Of course, experiments are central to science and the scientific process. From a historian's perspective, they might seem to be the most important elements, especially in resolving scientific disputes. But from a methodological perspective, experiments are one element in a process. Are for instance phenomena, observations, thought experiments, analogies, hypotheses, models, and theories not also important elements of the scientific process? Can they all be reduced to experiments, or would the result then be a naïve experimentalism? Mach (1905/1926/2002: 169) explicitly warns against abstracting experiments from their psychological process: "The Experiment guided by thought lies at the basis of science. It consciously and deliberately aims at widening experience. Nevertheless, one should not underrate the function of instinct and custom in the experiment. It is impossible to gain an instant intellectual survey of all the conditions that intervene in an experiment. […] In a field which has become familiar through continued concern with it, one goes about experiments quite differently."

The experiment as a scientific tool is thus strongly dependent on the previous experience. If the experience and a basic understanding of it do not exist, no new knowledge can result from any experiment. Without a theory (even if it is a naïve one), one cannot ask specific empirical questions regarding it, and the result will be white noise, or an answer not at all fitting to the theory. But in order to see this, one often needs to go back to the original phenomenon, not to its abstracted version in an experiment. As one can see here, experiments are much more metaphysical than the more empirical phenomena, and therefore stronger dependent on theory.

---

difference, the usage will be kept here. Also the term "*erkenntnis*-theoretical" will be used as an adverb.

[8] What is here called experimentalism is actually part of a wider phenomenon, which can be observed for several natural scientists, who tried to implement a Machian *Erkenntnistheorie* in science education, such as Henry Edward Armstrong in England, or Čeněk Strouhal in what is now the Czech Republic (details can be found in Siemsen 2010a and 2010b respectively). By training it is the instinct of the natural scientist to prefer the logical to the psychological approach. For Mach in strongly genetic processes, such as education, the psychological moment has to come before the logical one (Mach 1890). It should not be applied before it is developed, also not in the (meta)*erkenntnis*-theoretical analysis. Otherwise logic becomes an anthropomorphism (see Siemsen 2010a).

A theory therefore cannot be exclusively taught by experimentalism, which should only gain an increasing role as one becomes familiar with a field.[9] For instance, there is always a thought experiment preceding an experiment. When the result of the thought experiment is not clear, it seems necessary to do a physical experiment.[10] Without the mental process of going through the thought experiment resulting in an – equally mental – tension, the experimenter will not see any sense in a specific experiment.

Thus, for understanding the role of experiments in science, one needs to understand its whole psychological process. But at this point, Sarton shifts his attention away from the psychological research to the historical (i.e., to a specific version of the historical research, which largely abstracts from psychology versus a version in which the psychological questions are central for historical research). By this, he introduces an epistemological wedge between psychology and history, which does not exist for Mach. In order to have a closer look at this difference and the reasons for this difference, one needs to take a closer look at Mach's concepts of genesis and psychology of science.

Mach's psychology of science and research

Concerning *Erkenntnistheorie,* there have been two fundamental changes in modern science relative to the two-thousand-year-old Aristotelian system: the Copernican and the Darwinian revolution. The nature of these changes has become apparent only over time, and even now it is often overlooked or misunderstood. The Copernican and the Darwinian revolutions have changed the *erkenntnis*-theoretical nature of knowledge, although neither Copernicus nor Darwin was aware of all such consequences of their initial ideas. The Copernican world view finally shows (in its Machian, i.e., post-Newtonian interpretation) that empirically, one cannot define an absolute reference system, also not by defining any such system as "absolute" or

[9] This genetic problem in principle is important in all areas which are taught "new", i.e., for which there are no analogies to existing areas of knowledge with already laid empirical meanings. It therefore becomes even more important to introduce experimentalism to schools. Children cannot be assumed to already be small scientists in the sense that they can be expected to have full experience in any field of science. They intuitively use the method which is basic for scientific experimentation, i.e., the method of variation. Their use of it is though still too far away from scientific experimentation in order to jump over the difference. The bridging of this difference requires an increasing understanding of the whole scientific process on the one hand and a minimum of experiences in any scientific field on the other. Only then can the method be related to an intuitivized meaningful empirical basis. Mach describes this in terms of a general principle of teaching (1866: 2–3): "Once a part of science belongs to the literature, a second task remains, which is to popularize it, if possible. This second task also has its importance, but it is a difficult one. It has its importance, because – regardless of the distribution of knowledge that increases its value – it is not unimportant either for the further development of science itself how much knowledge has been disseminated into the public. The difficulty is to know the soil very well in which one wants to plant the knowledge. It is a prevalent but wrong opinion that children are not able to form precise concepts and come to the right conclusions. The child is often more sensible than the teacher. The child is very well able to comprehend, if one does not offer too much new at a time, but properly connects the new to the old. The adult is a child when facing the completely new. Even the scholar is a child when confronted with a foreign subject. The child is a child everywhere, as everything is new to him. The art of popularization lies in avoiding too much of the new at one time."

[10] Some of the greatest experiments are actually those, where from one gestalt perspective the result seems clear, but from another it is doubtful. The problem is to find another perspective, or better to start searching for it in the first place.

"god's thoughts". The "final jurisdiction" over any scientific world view is empirical, not metaphysical.[11] Similarly, Darwin's detailed synthesis of biological and geological facts exposed teleology as an anthropomorphic principle (see for instance Haeckel 1905). It was Mach who first brought the implications of the two together.[12] He showed in a synthesis the psychological and *erkenntnis*-theoretical shift necessary for a less anthropomorphic and consistent world view, which would integrate physics, physiology, psychology, mathematics and epistemology.[13]

Two empirical areas made an adaptation of the Aristotelian idea of science necessary: The facts observed for the "translunar" physics and the large number of fossils and geological as well as biological facts, which became apparent in the 17th, 18th and 19th centuries. There are two implications regarding scientific knowledge in which these facts are fundamentally inconsistent with the metaphysics of Aristotle: His initial *(a priori)* assumptions are partly inconsistent with the facts, and his "final cause" is teleological, and thus fundamentally anthropomorphic. Certainly, scientific thought will always contain many anthropomorphisms as the comparison to ourselves is too intuitive to be excluded, but all these anthropomorphisms have to be questionable, and none of them defined as final. Otherwise, scientific inquiry ends at a predefined point without empirical justification. Historical research shows how unjustified such views often were and suggest the humble assumption that our current views might be submitted to a similar verdict of the historians living a few centuries hence.

As a result, the empirical facts from all sciences have to be integrated into any general understanding of science, which certainly was the goal of Sarton in the first place.[14] This general idea of human (and scientific) knowledge can only be made fundamentally consistent if science is seen as a continuation (in terms of genesis) of general human thought. Science needs to be seen as fundamentally "one", i.e., there is a reason why we give all sciences the name "science", and this reason is fundamentally more important than any (conventional) categorization into different "sciences". This unifying idea of science must necessarily also encompass our epistemology and our world view, which are fundamentally dependent on our view of the physical and the psychical, i.e., psychophysics. As a result, no science can do without a "theory" of the mental realm. One can only disregard more elaborate thoughts about the topic, and use naïve psychology as a default (which several sciences at least in their mainstream movements have actually done). As a result, one introduces an arbitrary epistemological wedge, which easily leads to metaphysical inconsistencies, not only in-between sciences,

[11] This does not exclude that there are other areas of jurisdiction, but then the claims cannot be regarded as scientific claims.

[12] He published his initial ideas already in 1863, four years after the publication of Darwin's *Origin,* and much before Haeckel and Darwin himself tried to elaborate the implications of Darwin's theory for human (and scientific) knowledge.

[13] Mach (1890) stated that he had actually developed this view because of his research in psychophysics so that he would not constantly have to change his world view. Therefore he had been looking for a "neutral" view consistent with the physical, physiological, and psychical facts. He also stated that fundamentally one can only have one world view (as a fundamental gestalt). Seemingly different world views are therefore the results of inconsistent conceptual hierarchies.

[14] The epistemologically more elaborate version of this idea is monism, i.e., the idea that on the fundamental level "everything" should be "one". In Mach's interpretation, this is a requirement of an economy of thought, as one idea is far more economical for replicating thought than two ideas if one considers that all other ideas are derived from these. Thus, in the case of dualism, all successive ideas, concepts, etc. in principle need to be duplicated.

but even within. From this perspective many puzzling questions can only be regarded as pseudo-questions. Unfortunately, fundamental changes in world view require so many changes in the metaphysical and even empirical meanings of so many intuitivized concepts[15] that such fundamental gestalt changes are psychologically difficult and rare, especially among trained scholars. An increasingly complicated system of epicycles is (once learnt) easier to bear than the "*horror vacui*". Small, continuous gestalt adaptations require less mental investment in the short term, even if it might be very uneconomical in the long run.

Erkenntnistheorie and the "metapsychical"

The area of science concerning the questions above (from a Machian perspective) is in German called *Erkenntnistheorie*. In English, this area is often called "philosophy of science", although this is probably a misnomer. It suggests that it is a branch of philosophy, while actually its starting point is not philosophy with its traditional emphasis of metaphysical abstraction, but empirically oriented psychophysics (specifically including psychology).[16] As a result, what is today researched as "philosophy of science" is from a Machian perspective only a small part of what could be researched.

It is exactly this view of *Erkenntnistheorie* which has been the most Machian of Sarton's scientific ideas. These ideas also constitute Sarton's most original and most influential concepts. Unfortunately, they have been often misunderstood, and implemented with inconsistencies, i.e., arbitrary epistemological wedges, cutting-off the less understood parts of the ideas (often the Machian world view). As one could observe in the quote before, already Sarton himself introduced the first wedge by not knowing and not understanding Mach's psychology.[17]

[15] Concepts (also in science) become increasingly intuitive and thus less accessible to our conscious reflection with their use. Fundamental concepts are indirectly involved when thinking about higher-level concepts. Therefore the intuitivizing effect is exponential for very basic concepts, for instance concerning our world view. This is not necessarily corrected by empirical observations. As William James (1905/1967: 206) had observed, "I speak also of ideas which we might verify if we would take the trouble, but which we hold for true although unterminated perceptually, because nothing says 'no' to us, and there is no contradicting truth in sight. *To continue thinking unchallenged is, ninety-nine times out of a hundred, our practical substitute for knowing in the completed sense.* As each experience runs by cognitive transition into the next one, and we nowhere feel a collision with what we elsewhere count as truth or fact, we commit ourselves to the current as if the port were sure." So 99% of the times we just do not experience anything contradictory and this is normally interpreted as empirical agreement, though it might just be a sign of metaphysical "metastasising".

[16] Sarton seems to be aware of this empirical problem of world views when he states that the previous courses devoted to this have with the exception of Mach been far too philosophical. He thus makes a distinction between philosophy and *Erkenntnistheorie* (which he calls method).

[17] The problem of this misunderstanding is probably manifold, and has similarly occurred to many others close to Mach's thought. It is partly related to the editions of Mach's most philosophical work *Knowledge and Error* (1905): First, Sarton had read only the French translation of Mach's *Knowledge and Error* from 1908, which unfortunately is regarded as a very bad and fragmentary translation, which for instance omits most of the footnotes. The footnotes in Mach's works often contain the most important information. Secondly, Mach's publisher Paul Carus died before he could publish the book in English. The English translation appeared late, in 1976 (with a reprint of the title page of Sarton's French version at Harvard). Both, this English translation and the translation of Mach's previous *Analysis of Sensations* (and its predecessor, the *Contributions to the Analysis of Sensations*)

What is Mach's psychology, and especially his concept of "psychology of science"? Mach's psychology can be described with the help of four fundamental concepts (see also Siemsen 2010a). The first and foremost concept promoted by Mach himself is his concept of sensation, which he uses in the broadest sense one can probably imagine (Mach 1914). Mach's sensual elements comprise the whole psychophysical relation, e.g., from the sun's light via the visible body to the retina and including the physiological and cultural interpretation (abstractions) of what we actually see (for instance abstracting the physical body we actually see, into an "object"). For him, one cannot introduce any epistemologically consistent wedge anywhere in-between. It would "cut right through consciousness" (Mach 1905).

Secondly, Mach is the founding father of the Gestalt concept in psychology. This concept was adapted from Mach as an independent psychological concept by Christian von Ehrenfels, and further developed by Wertheimer, Köhler, etc. (see Siemsen 2010a for detailed quotations). Thirdly, Mach also is the founding father of Richard Semon's concept of *mneme* (Semon 1911, 1923).[18] With this concept, Gestalts can be genetically analyzed according to their current and their memorized elements.

Finally, Mach pioneered the concept of what I have called the "metapsychical". Just as sensations (i.e., in the Machian sense the adaptation of the thoughts to the facts) have physical, physiological, and psychical properties, theory (i.e., in the Machian sense the adaptation of the thoughts to each other) has metaphysical and "metapsychical" properties. These properties are initially dependent on the empirical physical, physiological, and psychical properties, but with increasing abstraction the empirical part is withered away. The metaphysical develops the internal consistency of ideas depending on the physical aspects of the facts (e.g., the internal logic of a theory); the metapsychical develops the ideas depending more on the psychical aspects of the facts (i.e., the method of thought developed through experience). A metapsychical analysis is necessary in order to distinguish the processes of adaptation of the thoughts to the facts and adaptation of the thoughts to each other, i.e., distinguishing empiry and metaphysics from each other. As the two are often intractably intertwined, especially in the origins, such an analysis is necessary for making the roots of scientific processes visible (which can certainly be considered a joint goal of Mach and Sarton) and thereby adaptable. The metapsychical analysis can thus be understood as the "unfolding"[19] of unconscious (intuitive) layers of lower-level gestalts. In this "unfolding" process, some layers may start to appear superfluous or arbitrary, which can then lead to further enquiries about possible manifolds.[20] This is what Mach means by his concept of "genesis".

---

cannot compare with the ingenious translation of Mach's *Mechanics* by Thomas J. McCormack. The problem translating Mach is that he changed his world view more than once with many interdependent conceptual gestalts shifting their meanings as well (like previously described for *Erkenntnistheorie*). As a result, even in German, Mach changes concepts and their meanings from edition to edition of his books. He continuously adapts old higher-level concepts to their new (internally consistent) meanings. This psychological process takes – as he noted – time and mental effort, a process which he never finished.

[18] This concept was picked up by Russell (1921/1922) and then reinvented by Dawkins (1976/1978) as "meme". The original concept by Semon is different and much more Machian though (see Siemsen 2010a for details).

[19] The metaphor of "folding" here can be understood similar to an origami figure, which assumes a final gestalt very different from its original form (i.e., a sheet of paper).

[20] Thus, the concept of the "metapsychical" does not denote a "metapsychological" theory, but the psychological process of the development of thought and scientific thought as a specific part of it. Nevertheless, the "metapsychical" does include aspects of a metapsychological theory, but with a specific, *erkenntnis*-theoretical focus.

An example: Sarton, Conant, experimentalism, and genesis

As an example, I will here briefly describe the metapsychical process for Sarton's influence on James Conant. Conant wrote in his obituary on Sarton (ISIS 1957b, George Sarton Memorial Issue) that

> [I] first met Sarton in the winter of 1916–1917 when I was beginning my academic career. […] I recall his saying emphatically that the history of science could not be written by historians – one must first of all have had a scientific training. […] During the evening's talk I fancied myself deserting chemistry and following the master's footsteps, but the fancy was short-lived. However, I continued to read what Sarton wrote and his ideal of a history in which Kings and their counselors as well as the battles were footnotes stayed with me for many years. His conviction as to the importance of the history of science certainly influenced my reading, my thinking and many years later, my writing. Therefore, I am indebted to Professor Sarton in a very personal sense.

If one reads and analyzes Conant's "later writings", one does not find many (if any) quotations on Sarton. Sarton's influence became so intuitive that Conant seemingly becomes aware of it only at the occasion of conscious reflection for the obituary article.[21] The strength of the influence also indicates that Conant must have intuitively understood the "deeper reasons" of the scientist to give his attention to the history of science, at least partly. Thereby, Conant had unknowingly acquired a part of Mach's *Erkenntnistheorie* and world view into his thinking and writing. He also intuitively adopted Sarton's deviations from Mach's *Erkenntnistheorie,* for instance the experimentalism. Conant's "Harvard Case Studies" for instance are explicitly about experimental science, although in his introduction he states (Conant 1957a: vii) that it is about "'understanding science'; it is independent of a knowledge of the scientific facts or techniques in the new area. [One should have a] 'feel' for what we may call 'the tactics and strategy of science.'" What Conant describes in the foreword is in principle Sarton's method or Machian *Erkenntnistheorie.*

From this perspective, Conant (1957a: viii/ix) notices the genetic problem and the cognitive gap between the learner and science up to a point:

> Modern science has become so complicated that today methods of research cannot be studied by looking over the shoulder of the scientist at work. If one could transport a visitor, however, to a laboratory where significant results were being obtained at an early stage in the history of a particular science, the situation would be far different. For when a science

[21] From the quotation, it becomes nevertheless clear that Conant describes this influence from Sarton as something quite deeply felt and not only as a matter of courtesy. In a metaphor, Einstein in his obituary to Mach (1916: 102/103) described that all the eminent physicists of his generation had been strongly influenced by Mach (mainly through the *Mechanics* and through Mach's widely used physic's school books), even though they even might not want to remember it: "I think that even those who think of themselves as enemies of Mach, don't remember how much of Mach's approach they have – so to speak – imbibed with their mother's milk." Similar to Conant's description of Sarton's influence on him, Einstein's metaphor implies that the earlier one drinks it, the more it will influence one's thinking, but the less one will remember how it came to this influence. It is another metapsychical phenomenon.

> is in its infancy, and a new field is opened by a great pioneer, the relevant information of the past can be summed up in a relatively brief compass. Indeed, if the methods of experimental science are being applied for the first time to a problem of importance, the scientist's knowledge would not be much greater than that of his inquiring guest. Briefly, and in simple language, he could explain the new experiment. Then as from day to day results were obtained and further experiments planned the visitor would see unfolding before him a new field. […] he had had a unique opportunity of learning at first hand about the methods of science.

Thus, though the genetic problem is in the centre, at the end the meaning of research is related mainly to experiments (which are, as shown before, largely metaphysical) instead of a continuous recourse to more empirical experience. Thereby a specific and genetically later part of the scientific process is taken as genetically early and general. If the visitor would have looked Einstein over the shoulder while he developed his ideas, he would have seen nothing – maybe his desk at the patent office, but not in terms of lab experiments. Similarly, the visitor would not have been able to observe Kekulé's benzene ring in the fireplace (or wherever he first thought about it), etc.

The question is what the visitor in Conant's thought experiment would really learn about the methods of science if he does not have any background knowledge, such as for instance knowing what and how to observe what Conant describes as a phenomenon in the first place. Conant certainly assumes a psychology of the visitor. But to what extent does the actual psychology of such a visitor follow this assumed psychology? Does it only do so (and only approximately) for the highly selected students of Harvard? To what extent is it generalizable?

One can quite certainly assume that if one uses Conant's own case studies, any average "layman" would be lost.[22] He would not have a "feel" for what chemistry is in the first place, what elements or chemical reactions are, nor would he have a concept of "air" anywhere close to how the chemists would use it in the description of the cases. For this, the empirical as well as the theoretical foundations are missing. Basic concepts are used before they have acquired any empirical meaning, any "feeling". As Mach (1890: 4) observed, "[In education,] criticism cannot begin where empirical meanings *[konkrete Vorstellungen]* are still lacking." In order to avoid this, one would have to go back, as Mach did, to the genetic origins of *Culture and Mechanics* (*Kultur und Mechanik,* Mach 1915). For the lay people, all of science, even the most basic concepts, have to be considered "new" in the sense of lacking prior empirical experiences. This concerns all scientific and pre-scientific experiences and not only those related to experimentation.[23], [24]

[22] The case studies start their in-depth analysis in the 16th–17th centuries the earliest.

[23] As John Bradley, a chemist strongly influenced by Mach, and a critic of the Nuffield experimental approach, recognized in his retirement speech (1975: 9): "Why have teachers […], including myself, failed so miserably? […] We have answered the question: Where does theory begin? wrongly. […] So with good intentions, we have said to Robert: What matters is the atom, or the molecule or the equation. Poor Robert has been stranded; he resembles a child aged six given logarithms to multiply three by two or like David he is too small and weak to carry the armor of Mendeleeff and Cannizzaro [both eminent chemists, the concepts of whom Bradley had earlier suggested to teach to students]. I am convinced that almost all of us have answered the question wrongly. Where has been our mistake? We have forgotten that all thought is theory, and that classification is thought and therefore also theory."

[24] It is interesting to note here that around the question of experience and experiment, different educationalists

In this sense, Sarton's critique on Mach's mechanics as not going far enough was right: Also Mach's *Mechanics* starts already from empirical and theoretical foundations, which need to have been laid (these may pre-exist in the excellent student, but not in the average one). Mach had this insight seemingly late in his life and just before his death published a book on *Prehistorical Inventions and the Origins of Mechanical Experience and Insight*. Unfortunately, the book is fragmentary, and not translated into English. It is a genetic addendum to Mach's *Mechanics,* based mainly on the observation and the experience of his son. Already in the introduction to his *Mechanics,* Mach had assumed that the knowledge of mechanics was adapted from the experiential treasures of the handcrafts by intellectual refining. *Culture and Mechanics* starts genetically much earlier and integrates general culture better, although the conceptualization of what is mechanical from the totality of sensual elements becomes necessarily arbitrary. Mach in his foreword sees the book only as the beginning of a research in the origins of mechanics, and maybe a founding of a "general genetic technology". Here Mach seems close to Sarton's late project of writing the history of "Ancient Science". The difference is mainly that Sarton is limited in his genetic approach to historic times, while Mach with his general genetic method goes back further into "pre-historic" times of the origins of science. But both seemingly pursue the same idea at the end of their lives.

In this joint direction, Sarton later even developed an intuition about the psychology of fundamental Gestalt switches in science (without calling them Gestalts though): "When such a point has been reached it is generally possible to make enormous steps forward [...] It is almost as if we had traversed a range of insurmountable mountains by means of a tunnel and found ourselves for the first time on the other side beholding a landscape never seen before" (George Sarton 1931).[25]

References

Bradley, J. (1975) Where does theory begin? *Education in Chemistry* 1975 (March), 8–11.

Conant, J. B. (ed.) (1957a) *Harvard Case Histories in Experimental Science. Vol. 1.* Cambridge: Harvard University Press.

Conant, J. B. (1957b) George Sarton and Harvard University. *ISIS* 48 (3), 301–305.

Dawkins, R. (1976/1978) *The Selfish Gene.* New York: Oxford University Press.

Einstein, A. (1916) Ernst Mach. *Physikalische Zeitschrift* 17 (7, April 1), 101–104.

Haeckel, E. (1905) Der Kampf um den Entwicklungs-Gedanken: Drei Vorträge, gehalten am 14., 16. und 19. April 1905 im Saale der Sing-Akademie zu Berlin. Berlin: Reimer.

Holton, G. (2003) The Project Physics Course, Then and Now. *Science & Education* 12 (8), 779–786.

James, W. (1905/1967/1977) A World of Pure Experience. In: McDermott, J. J. (Ed.), *The Writings of William James – A Comprehensive Edition* (194-213). Chicago: University of Chicago Press.

close to Mach's ideas have taken quite diverging directions. For instance, Armstrong favored the "heuristic" approach, i.e., in principle based on experience, but strongly biased towards experiments, taking historical questions little into account. Martin Wagenschein, a German physics educator, favored natural phenomena, but rejected the very idea of statistically measuring results from educational experiments.

[25] In a metapsychical analysis, Sarton metaphorically describes the psychical phenomenon of a gestalt shift, although he is certainly not consciously aware of this fact. The idea of the metapsychical analysis only makes this a deliberate scientific undertaking instead of a tacitly implied one.

Kuhn, T. S. (1962) *The Structure of Scientific Revolutions.* Chicago: University of Chicago Press.
Mach, E. (1866) Einleitung in die Helmholtz'sche Musiktheorie – Populär für Musiker dargestellt. Graz: Leuschner & Lubensky.
Mach, E. (1883/1933/1976) *Die Mechanik – historisch-kritisch dargestellt.* Reprint of the 9th edition. Darmstadt: Wissenschaftliche Buchgesellschaft.
Mach, E. (1890) Über das psychologische und logische Moment im Naturwissenschaftlichen Unterricht. *Zeitschrift für den physikalischen und chemischen Unterricht* 4 (1, October 1890), 1–5.
Mach, E. (1893/1960) The Science of Mechanics: A Critical and Historical Account of Its Development. La Salle: Open Court.
Mach, E. (1905/1926/2002): *Erkenntnis und Irrtum: Skizzen zur Psychologie der Forschung.* 5th edition. Leipzig, Berlin: Reprint by rePRINT.
Mach, E. (1914) The Analysis of Sensations and the Relation of the Physical to the Psychical. Chicago: Open Court.
Mach, E. (1915) *Kultur und Mechanik.* Stuttgart: von W. Spemann.
de Mey, M. (1984) George Sarton's Concept of Science Studies at Ghent during His Time and in Ours. In: W. Callebaut et al. *George Sarton Centennial.* Communication & Cognition, 3–6.
Russell, B. (1921/1922) *The Analysis of Mind.* London: Allen & Unwin.
Sarton, G. (1916) The history of science. *The Monist* 26: 321–365.
Sarton, G. (1918) The teaching of the history of science. *The Scientific Monthly* 7 (3), 193–211.
Sarton, G. (1931/1962) The history of science and the history of civilization. In: *The History of Science and the New Humanism.* Bloomington: Indiana University Press.
Sarton, G. (1952/1970) A History of Science: Ancient Science Through the Golden Age of Greece. New York: The Norton Library.
Sarton, G. (1957) *Six Wings: Men of Science in the Renaissance.* Bloomington: Indiana University Press.
Sarton, G. (1959) Hellenistic Science and Culture in the Last Three Centuries B.C. Cambridge: Harvard University Press.
Semon, R. (1911/1920) Die Mneme als erhaltendes Prinzip im Wechsel des organischen Geschehens. Leipzig: Engelmann.
Semon, R. (1923) *Mnemic Psychology.* London: Allen & Unwin.
Siemsen, H. (2010a – forthcoming) Intuition in the scientific process and the intuitive "error" of science. In: F. Columbus (ed.) *The Psychology of Intuition.* Hauppauge: Nova Science.
Siemsen, H. (2010b – in print) Mach's science education, the PISA study and Czech science education. In: A. Mizerova (ed.) *Ernst Mach: Fyzika–Filosofie–Vzdělávání.* Brno: Masaryk University Press.
Siemsen H., Siemsen, K. H. (2009) Resettling the Thoughts of Ernst Mach and the Vienna Circle to Europe – The cases of Finland and Germany. *Science & Education* 18 (3): 299-323.
Wertheimer, M. (1924/1938) Gestalt theory. In: W. D. Ellis (ed.) *A Source Book of Gestalt Psychology* (1–11). London: Kegan, Trench, and Trubner.